



(51) International Patent Classification:
C40B 30/04 (2006.01) *C12Q 1/68* (2006.01)

(21) International Application Number:
PCT/US2011/031450

(22) International Filing Date:
6 April 2011 (06.04.2011)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/321,172 6 April 2010 (06.04.2010) US

(71) Applicant (for all designated States except US): **THE GEORGE WASHINGTON UNIVERSITY** [US/US]; 2300 Eye Street, N.W., Suite 712, Washington, DC 20037 (US).

(72) Inventor; and

(75) Inventor/Applicant (for US only): **WU, Valerie, Wailin** [US/US]; 16610 Leopard Terrace, Rockville, MD 20854 (US).

(74) Agent: **SCHOFIELD, Mary Anne**; Crowell & Moring LLP, Intellectual Property Group, P.O. Box 14300, Washington, DC 20044-4300 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM,

AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

- with international search report (Art. 21(3))
- before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))
- with sequence listing part of description (Rule 5.2(a))

(54) Title: COMPOSITIONS AND METHODS FOR IDENTIFYING AUTISM SPECTRUM DISORDERS

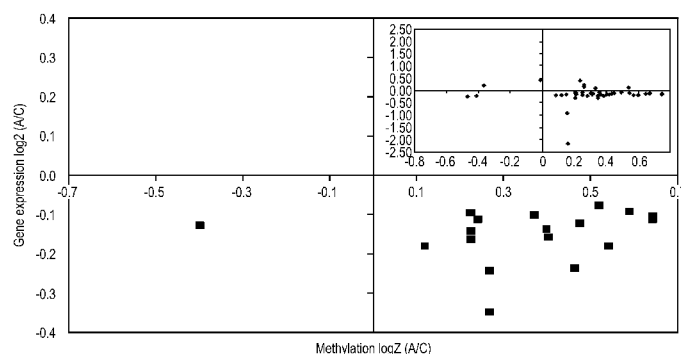


FIG. 1A

(57) Abstract: The invention provides methods of identifying DNA methylation profiles for neurological and psychiatric conditions including autism spectrum disorders, methods of treating such conditions, and methods of identifying therapeutics for the treatment of such neurological and psychiatric conditions.

COMPOSITIONS AND METHODS FOR IDENTIFYING AUTISM SPECTRUM DISORDERS

FIELD OF THE INVENTION

[0001] This invention relates to DNA methylation microarray technology, and more specifically to methods and kits for identifying autism and autism spectrum disorders in humans.

BACKGROUND OF THE INVENTION

[0002] Autism spectrum disorders (ASD) collectively represent a heterogeneous range of pervasive neurodevelopmental disorders that are characterized by deficits in social interactions and understanding, aberrant communication and/or language development, and restricted interests and stereotyped behaviors (Volkmar, F. R. et al (1994) *Am J Psychiatry* 151, 1361-71; Siegel, B. (1991) *Psychiatr Clin North Am* 14, 53-68). It is believed that multiple etiologies contribute to the heterogeneity in clinical presentation of ASD and further complicate efforts to identify genetic/epigenetic or molecular markers of the disease.

[0003] While autism is largely considered genetic in origin due to its high heritability observed in twin and family studies (Bailey, A. et al. (1995) *Psychol Med* 25, 63-77; Folstein, S., and Rutter, M. (1977) *J Child Psychol Psychiatry* 18, 297-321), recent evidence supports the involvement of epigenetic regulatory mechanisms in the pathogenesis of ASD (Schanen, N. C. (2006) *Hum Mol Genet* 15 Spec No 2, R138-50; Beaudet, A. L. (2007) *Nature Medicine* 13, 534-536; Nakayama, A. et al (2006) *Nihon Shinkei Seishin Yakurigaku Zasshi* 26, 209-12). These regulatory mechanisms include DNA methylation at CpG sites, genomic imprinting, chromatin modifications, and non-coding RNA (Mehler, M. F., and Mattick, J. S. (2006) *J Physiol* 575, 333-41; Champagne, F. A. (2008) *Frontiers in Neuroendocrinology* 29, 386-397; Chuang, J. C., and Jones, P. A. (2007) *Pediatr Res* 61, 24R-29R; van Vliet, J., Oates, N. A., and Whitelaw, E. (2007) *Cell Mol Life Sci* 64, 1531-8; Kaminsky, Z. et al. (2006) *Ann Med* 38, 530-44; Lopez-Rangel, E., and Lewis, M. E. (2006) *Clin Genet* 69, 21-2). Two single-gene disorders that may be associated with autistic symptoms, Fragile X Syndrome and Rett's Syndrome, arise from epigenetic dysregulation. Fragile X syndrome results from an expansion of the trinucleotide CGG repeat in the 5'UTR region of the FMR1 gene, leading to increased susceptibility to methylation and epigenetic silencing of FMR1 (Feng, Y. et al. (1995) *Science* 268, 731-4; Li, Z. et al., (2001) *Nucleic Acids Res* 29, 2276-83), and Rett's Syndrome arises from a mutation in the methyl CpG binding protein 2 (MeCP2) gene, whose protein product is responsible for recognizing methylated genes and is one of the key mediators of epigenetic regulation (Van Den Veyver, I. B., and Zoghbi, H. Y. (2001) *Brain and Development* 23, S147-S151; Amir, R. E. et al. (1999) *Nat Genet* 23, 185-8).

[0004] Hu *et al.* recently demonstrated differential gene expression in lymphoblastoid cell lines (LCL) from monozygotic twins discordant for diagnosis of autism (Hu, V. et al. (2006) BMC Genomics 7, 118), which strongly suggests that epigenetic factors are also involved in *idiopathic* autism. Other studies have suggested that “epigenetic hotspots” or regions susceptible to genomic imprinting are located in chromosomal regions (e.g., 15q and 7q) identified in genetic linkage analysis of autism (Schanen, N. C. (2006) Hum Mol Genet 15 Spec No 2, R138-50; Davies, W. et al. (2001) Ann Med 33, 428-36). Hogart et al. (Hogart, A. et al. (2007) Hum Mol Genet 16, 691-703) argues that genes located close to these hotspots (like genes encoding for GABAA-receptor subunits, GABRB3, GABRA5 and GABRG3), while not necessarily subject to imprinting, can still convey an ASD risk upon disrupted epigenetic regulation.

[0005] Methylation of cytosine residues of CpG sites is the most characterized of the epigenetic mechanisms and involves the addition of a methyl group onto the 5' position of a cytosine residue. While 60-90% of CpGs are found methylated throughout the genome (Klose and Bird, A. P. (2006) Trends in Biochemical Sciences 31, 89-97) unmethylated CpG sites can be found clustered in CpG islands most often associated with 5' promoter regions of genes. Methylation of CpG islands is most often associated with transcriptional silencing and has been found to be a significant contributor to altered gene expression. The mechanism of methylation as a silencing signal is thought to occur by either recruitment of repressive transcriptional silencing machinery, or by steric hindrance preventing the binding of transcription factors necessary for transcriptional activation.

[0006] Thus, there is a need for compositions and methods that will provide an increased understanding of the pathophysiology of Autism spectrum disorders, such as autism, pervasive developmental disorders not otherwise specified (PDD-NOS), and Asperger's syndrome, and their treatment.

[0007] The present invention satisfied these and other needs by demonstrating herein that one of the regulatory mechanisms responsible for altered gene expression in ASD is altered methylation status of genes, that is, differential methylation of genes relative to that of control samples from non-autistic individuals. The present invention provides compositions and methods for DNA methylation differential gene expression profiling and reveals significantly differentially DNA-methylated genes that are associated with neurological diseases, nervous system development and function, as well as other co-morbid disorders associated with ASD, such as gastrointestinal, muscular, and inflammatory disorders.

SUMMARY OF THE INVENTION

[0008] One aspect of the invention provides an array, e.g., a gene chip array or focussed array, having a plurality of different oligonucleotides with specificity for differentially methylated DNA gene promoter regions, or CpG islands (regions of DNA containing relatively high content of CpG dinucleotides), of genes that are associated with at least one autism spectrum disorder, wherein the autism spectrum disorder comprises autistic disorder, pervasive developmental disorder-not otherwise specified (PDD-NOS), including atypical autism, Asperger's Disorder, or a combination thereof.

[0009] In one embodiment of the present invention, an oligonucleotide array, e.g., gene chip array, is provided wherein the oligonucleotides are specific for at least one of the genes set out in Table 1, Table 2, or Table 3, or a combination thereof. The oligonucleotides are preferably specific for unmethylated or methylated regions in the gene(s) or for unmethylated or methylated CpG islands of the promoter(s) of the gene(s).

[0010] In another aspect of the invention, a method of screening a subject for a neurological disease or disorder comprising the steps of: (a) isolating a sample of nucleic acid, protein or cellular extract from at least one cell from the subject; (b) measuring the level of DNA methylation of at least one gene in Table 1, Table 2, Table 3 or a combination thereof in a sample, wherein the at least one gene has been determined to have differential DNA methylation pattern in subjects with a neurological disease or disorder as compared to a control sample from individuals without the neurological condition, wherein the subject is diagnosed to be at risk for or affected by a neurological disease or disorder if there is a statistically significant difference in the level of DNA methylation of at least one gene listed in Table 1, Table 2, or Table 3 compared to the level of DNA methylation of the same genes from a sample from healthy individuals without the neurological condition.

[0011] In one embodiment of the screening method of the present invention, the neurological disease comprises at least one autism spectrum disorder, autistic disorder, pervasive developmental disorder-not otherwise specified (PDD-NOS) including atypical autism, Asperger's Disorder, or a combination thereof.

[0012] In another embodiment of the screening method of the present invention, the at least one differentially DNA methylated gene in Table 1, Table 2, Table 3 or a combination thereof comprise genes involved in nervous system development, axon guidance, synaptic transmission or plasticity, myelination, long-term potentiation, neuron toxicity, Purkinje cell differentiation, cerebellar development, embryonic development, regulation of actin networks, digestion, inflammation,

oxidative stress, epilepsy, apoptosis, morphogenesis, cell survival, differentiation, the unfolded protein response, Type II diabetes and insulin signaling, digestion, liver toxicity (hepatic stellate cell activation, fibrosis, and cholestasis), endocrine function, circadian rhythm, cholesterol metabolism and the steroidogenesis pathway, or a combination thereof.

[0013] In yet another embodiment of the screening method of the present invention, the healthy individual without the neurological condition is a non-phenotypic discordant twin, sibling of the subject, or healthy, unrelated individual.

[0014] In yet another embodiment of the screening method of the present invention, the method distinguishes between different variants of autism spectrum disorder comprising a lower severity scores across all ADIR items, an intermediate severity across all ADIR items, a higher severity scores on spoken language items on the ADIR, a higher frequency of savant skills, and a severe language impairment, or a combination thereof.

[0015] In yet another embodiment of the screening method of the present invention, the level of DNA methylation is quantified with an assay comprising bisulfite-sequencing, methylation-specific PCR (MSP), large scale CpG island microarray analysis using methylation-enriched and -unenriched samples, promoter analysis using methylation-enriched and -unenriched samples, combined bisulfite and restriction analysis (COBRA), pyrosequencing, Methyl-Light (Premier Biosoft International) or Methyl-Profiler (SABiosciences) methylation analysis, Sequenom mass analyzer, bisulfite treatment of DNA combined with selected amplification of the specific promoter regions of differentially methylated genes with the addition of a T7 promoter tag, followed by T7-mediated transcription and RNase T1 cleavage of the transcript with methylation sites determined by MALDI-TOF mass spectrometry analysis in the manner of Schatz et al. (2004) *Nucleic Acids Res.* 32(21): e167, or a combination thereof.

[0016] In yet another aspect of the invention, a method is provided for determining a differential DNA methylation profile for at least one autism spectrum disorder, comprising (a) preparing samples of control and experimental DNA, wherein the experimental DNA is generated from a nucleic acid sample isolated from a subject suspected of being afflicted with the at least one autism spectrum disorder and the control DNA is generated from a nucleic acid sample isolated from a healthy individual; (b) preparing one or more microarrays comprising a plurality of different oligonucleotides having specificity for differentially methylated genes associated with the at least one autism spectrum disorder; (c) applying the prepared samples to the one or more microarrays to allow hybridization between the oligonucleotides and the control DNA and the oligonucleotide and the experimental DNA ; (d) identifying the oligonucleotides on the microarray which display differential hybridization to the experimental DNA relative to the control DNA thereby determining a

differential DNA methylation profile for the at least one autism spectrum disorder.

[0017] In one embodiment of the differential DNA methylation profiling method of the present invention, the plurality of different oligonucleotides is specific for at least one differentially methylated gene in Table 1, Table 2, Table 3 or a combination thereof.

[0018] In another embodiment of the differential DNA methylation profiling method of the present invention, the at least one autism spectrum disorder comprises autistic disorder, pervasive developmental disorder-not otherwise specified (PDD-NOS), including atypical autism, Asperger's Disorder, or a combination thereof.

[0019] In yet another aspect of the invention, a method is provided for distinguishing between different phenotypes of an autism spectrum disorder comprising severely language impaired (L), mildly affected (M), or "savants" (S) comprising (a) preparing samples of control and experimental DNA, wherein the experimental DNA is generated from a nucleic acid sample isolated from a subject suspected of being afflicted with at least one phenotype comprising the severely language impaired (L), mildly affected (M), or "savants" (S); (b) preparing one or more microarrays comprising a plurality of different oligonucleotides having specificity for differentially methylated genes associated with the at least one phenotype; (c) applying the prepared samples to the one or more microarrays to allow hybridization between the oligonucleotides and the control and experimental DNAs; (d) identifying the oligonucleotides on the microarray which display differential hybridization to the experimental DNA relative to the control DNA thereby determining a DNA methylation profile for distinguishing among the different phenotypes of autism spectrum disorder.

[0020] In another embodiment of the phenotype distinguishing method of the present invention, the plurality of different oligonucleotides is specific for at least one differentially methylated gene in Table 1, Table 2, Table 3 or a combination thereof.

[0021] In yet another embodiment of the phenotype distinguishing method of the present invention, the at least one autism spectrum disorder comprises autistic disorder, pervasive developmental disorder-not otherwise specified (PDD-NOS), including atypical autism, Asperger's Disorder, or a combination thereof.

[0022] In yet another aspect of the invention, a method is provided for predicting efficacy of a test compound for altering a behavioral response in a subject with at least one autism spectrum disorder comprising: (a) preparing a microarray comprising a plurality of different oligonucleotides, wherein the oligonucleotides have specificity for differentially DNA-methylated gene promoter regions or CpG islands associated with at least one autism spectrum disorder; (b) obtaining a differential

DNA-methylation profile representative of the differential DNA methylation profile of at least one sample of a selected tissue type from a subject subjected to each of at least one of a plurality of selected behavioral therapies which promote the behavioral response; (c) administering the test compound to the subject; and (d) comparing a differential DNA methylation profile data in at least one sample of the selected tissue type from the subject treated with the test compound to determine a degree of similarity with one or more differential DNA methylation profile associated with an autism spectrum disorder; wherein the predicted efficacy of the test compound for altering the behavioral response is correlated to said degree of similarity.

[0023] In another embodiment of the compound efficacy testing method of the present invention, the plurality of oligonucleotides is specific for at least one differentially methylated gene in Table 1, Table 2, Table 3 or a combination thereof. The oligonucleotides may be specific for, e.g., (i) the promoter region of the gene, or CpG island(s) of the gene, as methylated in healthy individuals without a neurological disorder, e.g., an autism spectrum disorder, or (ii) specific for the promoter region of the genes or CpG island(s) as methylated in a subject having a neurological disorder, e.g., an autism spectrum disorder, or a combination of both (i) and (ii).

[0024] In yet another embodiment of the compound efficacy testing method of the present invention, the autism spectrum disorder neurological condition comprises autistic disorder, pervasive developmental disorder-not otherwise specified (PDD-NOS), including atypical autism, Asperger's Disorder, or a combination thereof.

[0025] In yet another embodiment of the compound efficacy testing method of the present invention, step (a) comprises obtaining a differential DNA methylation profile representative of the differential DNA methylation profile of at least two samples of a selected tissue type.

[0026] In yet another embodiment of the compound efficacy testing method of the present invention, the selected tissue type comprises a neuronal tissue type.

[0027] In yet another embodiment of the compound efficacy testing method of the present invention, the neuronal tissue type is selected from the group consisting of olfactory bulb cells, cerebrospinal fluid, hypothalamus, amygdala, pituitary, nervous system, brainstem, cerebellum, cortex, frontal cortex, hippocampus, striatum, and thalamus.

[0028] In yet another embodiment of the compound efficacy testing method of the present invention, the selected tissue type is selected from the group consisting of lymphocytes, blood, or mucosal epithelial cells, brain, spinal cord, heart, arteries, esophagus, stomach, small intestine, large intestine, liver, pancreas, lungs, kidney, urinary tract, ovaries, breasts, uterus, testis, penis,

colon, prostate, bone, muscle, cartilage, thyroid gland, adrenal gland, pituitary, bone marrow, blood, thymus, spleen, lymph nodes, skin, eye, ear, nose, teeth or tongue.

[0029] In yet another embodiment of the compound efficacy testing method of the present invention, the test compound is an antibody, a nucleic acid molecule, a small molecule drug, or a nutritional or herbal supplement.

[0030] In yet another embodiment of the compound efficacy testing method of the present invention, the behavioral therapy comprises applied behavior analysis (ABA) intervention methods, dietary changes, exercise, massage therapy, group therapy, talk therapy, play therapy, conditioning, or alternative therapies such as sensory integration and auditory integration therapies.

[0031] In yet another aspect of the invention a method is provided for assessing the efficacy of a treatment in an individual having at least one autism spectrum disorder comprising (a) determining the level of DNA methylation of at least one differentially DNA-methylated genes in Table 1, Table 2, Table 3 or a combination thereof in the sample, or a combination thereof, in a plurality of patient samples of a selected tissue type; (b) determining a degree of similarity between (a) the level of DNA methylation of at least one differentially DNA-methylated genes in Table 1, Table 2, Table 3 or a combination thereof in the sample profile data in the patient samples; and (b) the level of DNA methylation of at least one differentially DNA-methylated genes in Table 1, Table 2, Table 3 or a combination thereof in the sample profile specific for the genes set out in listed in Table 1, Table 2, Table 3, , or a combination thereof, produced by a therapy which has been shown to be efficacious in treatment of the at least one autism spectrum disorder; wherein a high degree of similarity of the differential methylation profile data is indicative that the treatment is effective.

[0032] In yet another embodiment of the invention a method is provided for assessing the efficacy of a treatment in an individual having at least one autism spectrum disorder comprising (a) determining a DNA methylation profile data specific for at least one differentially DNA-methylated genes set out in Table 1, Table 2, Table 3 or a combination thereof, in a plurality of patient samples of a selected tissue type; (b) determining a degree of similarity between (a) the DNA methylation profile data in the patient samples; and (b) a DNA methylation profile specific for the genes set out in listed in Table 1, Table 2, Table 3 or a combination thereof, produced by a therapy which has been shown to be efficacious in treatment of the at least one autism spectrum disorder; wherein a high degree of similarity of the DNA methylation profile data is indicative that the treatment is effective.

[0033] In yet another aspect of the invention, a method is provided for determining a DNA methylation profile indicative of administration of a therapeutic treatment to a subject with at least one autism spectrum disorder comprising (a) preparing samples of control and experimental

nucleic acid, wherein the experimental nucleic acid is generated from a nucleic acid sample isolated from a subject who has received the therapeutic treatment; (b) preparing one or more microarrays comprising a plurality of different oligonucleotides, wherein the oligonucleotides are specific for genes associated with an autism spectrum disorder; (c) applying the prepared samples to the one or more microarrays to allow hybridization between the oligonucleotides and the control and experimental nucleic acids; (d) identifying the oligonucleotides on the microarray which display differential hybridization to the experimental nucleic acid relative to the control nucleic acid thereby determining a differential DNA methylation profile indicative for the administration of the therapeutic treatment to the subject with at least one autism spectrum disorder.

[0034] In one embodiment of the method of the present invention, the plurality of different oligonucleotides is specific for at least one differentially methylated gene in Table 1, Table 2, Table 3, or a combination thereof.

[0035] In another embodiment of the method of the present invention, the at least one autism spectrum disorder neurological condition comprises autistic disorder, pervasive developmental disorder-not otherwise specified (PDD-NOS), including atypical autism, Asperger's Disorder, or a combination thereof.

[0036] In yet another aspect of the invention, a method is provided for conducting drug discovery comprising (a) generating a database of DNA methylation profile data representative of the genetic expression response of at least one selected neuronal tissue type from a subject that was subjected to at least one of a plurality of behavioral therapies and that has undergone a selected physiological change since commencement of the behavioral therapy; (b) administering small molecule test agents to untreated subjects to obtain DNA methylation profile data associated with administration of the agents and comparing the obtained data with the one or more selected DNA methylation profiles; (c) selecting test agents that induce DNA methylation profiles similar to DNA methylation profiles obtainable by administration of behavioral therapy; (d) conducting therapeutic profiling of the selected test compound(s), or analogs thereof, for efficacy and toxicity in subjects; and (e) identifying a pharmaceutical preparation including one or more agents identified in step (d) as having an acceptable therapeutic and/or toxicity profile.

[0037] In another embodiment of the method of the present invention, the behavioral therapy comprises applied behavior analysis (ABA) intervention methods, dietary changes, exercise, massage therapy, group therapy, talk therapy, play therapy, conditioning, or alternative therapies such as sensory integration and auditory integration therapies.

[0038] In yet another embodiment of the method of the present invention, the selected physiological change includes one or more improvements in social interaction, language abilities, restricted interests, repetitive behaviors, sleep disorders, seizures, gastrointestinal, hepatic, and mitochondrial function, neural inflammation, or a combination thereof.

[0039] In yet another embodiment of the method of the present invention, prior to administration of behavioral therapy, the subject shows at least one symptom of a psychological or physiological abnormality.

[0040] In yet another embodiment of the method of the present invention, the neuronal tissue type is selected from the group consisting of olfactory bulb cells, cerebrospinal fluid, hypothalamus, amygdala, pituitary, nervous system, brainstem, cerebellum, cortex, frontal cortex, hippocampus, striatum, and thalamus.

[0041] In one embodiment of the compositions of the present invention, the gene chip array having a plurality of different oligonucleotides with specificity for the differentially DNA-methylated gene promoter regions or CpG islands of retinoic acid receptor-related orphan receptor alpha (RORA), B cell CLL/lymphoma 2 gene (BCL-2), or a combination thereof.

[0042] In one embodiment of the compositions of the present invention, the gene chip array having a plurality of different oligonucleotides with specificity for the differentially DNA-methylated gene promoter regions or CpG islands one or more genes of Table 1, Table 2, Table 3, or a combination thereof, and specifically excludes a plurality of oligonucleotides with specificity the differentially DNA-methylated MeCP₂ gene and the oxytocin receptor gene, or a combination thereof.

[0043] In one embodiment of each of the methods of the present invention, the differentially methylated gene comprises RORA *or* BCL-2, or a combination thereof.

[0044] In one embodiment of each of the methods of the present invention, the differentially methylated gene specifically excludes the differentially DNA-methylated MeCP₂ gene and the oxytocin receptor gene, or a combination thereof.

[0045] In yet another aspect of the invention, a kit is provided for identifying a compound for treating at least one autism spectrum disorder comprising (a) a database having information stored therein one or more DNA methylation profiles specific for the genes set out in listed in Table 1, Table 2, Table 3 or a combination thereof, of subjects that have been subjected to at least one of a plurality of selected autism spectrum disorder neurological therapies and wherein the subject has undergone a desired physiological change; and (b) a computer program for comparing DNA

methylation profile data obtained from assays wherein a test compound is administered to a subject with the database and providing information representative of a measure of similarity between the DNA methylation profile data and one or more stored DNA methylation profiles.

[0046] In yet another aspect of the invention, a computer-readable medium on which is encoded programming code for analyzing autism spectrum disorder from a plurality of data points comprising a profile of differentially expressed DNA methylation genes, wherein said differential DNA methylation profile is specific for at least one differentially DNA methylated gene set out in Table 1, Table 2, or a combination thereof.

[0047] In yet another embodiment of the invention, in each of the screening methods, DNA methylation profiling methods, phenotype distinguishing methods, drug discovery methods, compound efficacy testing methods, computer program for determining a DNA methylation profile, and kits specifically provided for *supra* (and *infra*) may also be, without any limitation, made and/or practiced with at least two to at least fifty, or any integer value thereof, different differentially DNA methylated genes set out in Table 1, Table 2, Table 3, or a combination thereof.

BRIEF DESCRIPTION OF THE DRAWINGS

[0048] The foregoing and other aspects and advantages of the invention will be appreciated more fully from the following further description thereof, with reference to the accompanying drawings wherein:

[0049] Figure 1 depicts the overlap of dysregulated genes identified from CpG island microarray and global gene expression profiling previously performed using the same samples. A) Scatter plot shows inverse correlation of log2 ratios of gene expression and methylation of genes with differentially methylated CpG islands directly overlapping with its 5' end. The inset graph plots log2 ratios of genes with differentially methylated CpG islands located either upstream or downstream of the CpG island. Interestingly, the majority of genes identified from these analyses were hypermethylated in the twins relative to their respective nonautistic siblings, with a corresponding decrease in gene expression. B) Network analysis of 25 genes (circumscribed in dashed line) that were both differentially methylated and expressed. The network was generated using Pathway Studio 5 network prediction software (Ariadne Genomics, Rockville, MD) and identified common biological themes including apoptosis, cellular differentiation and inflammation. The analysis also revealed neurologically relevant functions and disorders including synaptic regulation, development, and mental deficiency.

[0050] Figure 2 depicts bisulfite sequencing of the BCL-2 P1 promoter region (shown in A). Bisulfite sequencing was performed across 38 CpG sites from genomic DNA isolated from LCL of discordant

MZ twins (B). The autistic co-twins are designated by "A_" preceding the blood identifier number (e.g., A_809), and the undiagnosed co-twins are designated by "M_" preceding the blood number (e.g., M_810). Control individuals (C) include unaffected siblings (e.g., C_813), one normal monozygotic twin pair (C_2744 and C_2745). A pair of siblings, one autistic (A_2020) and one unaffected (C_2019), was also included in this analysis. Inset table shows average % methylation with standard deviation (SD) across all 38 CpG sites for each group of samples.

[0051] Figure 3 depicts representative image of methylation specific-PCR (MSP) of the upstream CpG island region of RORA (shown in A) in LCL from autistic (A_737 and A_2020) and unaffected control (C_735 and C_2019) siblings. LCLs were treated with DMSO (vehicle control) or 5uM of 5-Aza-2-deoxycytidine for 48 hours. Genomic DNA was isolated and bisulfite modified. B) Modified DNA was used for PCR containing primers specific for unmethylated (U) and methylated (M) CpG sites of RORA.

[0052] Figure 4 depicts qRT-PCR of BCL-2 and RORA transcripts from LCL of 3 pairs of discordant twins and respective sibling controls for 2 pairs of twins. (A) Results are shown as percent of gene expression of autistic co-twin (#_A). (B) RNA isolated from LCLs following 48hr treatment with 5uM 5-Aza-2-deoxycytidine was also analyzed by qRT-PCR. Results shown are the average fold-change $\pm$ SE in expression following treatment.

[0053] Figure 5 depicts higher magnification (40x) images of immunohistochemical staining for RORA and BCL-2 on cerebellar sections from autistic and age-matched control subjects. ML: molecular layer; PL: Purkinje layer; GL: granule layer. Arrows point to Purkinje cells.

[0054] Figure 6 depicts RORA protein also appears to be reduced in the frontal cortex (BA9) in the majority of tissue sections from age- and sex-matched autistic and control subjects. This tissue array, obtained through the Autism Tissue Program, was prepared as described by Nagarajan et al. (2006) (Nagarajan, R. P. et al (2006) Epigenetics 1 172-182) in the laboratory of Dr. Janine LaSalle.

DETAILED DESCRIPTION OF THE INVENTION

[0055] The invention disclosed herein provides methods and compositions for diagnosis and treatment of neurological conditions. In particular, the invention provides microarray technology to diagnose and treat autism spectrum disorders. The invention relates, in part, to sets of genetic markers whose DNA methylation patterns correlate with therapeutic treatments of neurological, and in particular, autism spectrum disorders.

[0056] The invention provides not only methods of identifying differentially methylated gene profiles

for neurological conditions, but also methods of using such differentially methylated gene profiles in order to select particular therapeutic compounds useful in the prevention and treatment of such neurological conditions. The invention further relates to the application of differentially methylated gene profiles for the identification of therapeutic targets, and related pharmaceutical methods and kits.

[0057] The systems and methods described herein include microarray systems including gene chips and arrays of nucleotide sequences for detecting differentially methylated gene profiles of neurological conditions, and in particular, autism spectrum disorder conditions. The systems and methods described herein provide microarrays that have a plurality of oligonucleotide primers immobilized thereon and have specificity for differentially methylated genes associated with neurological conditions, and in particular, autism spectrum disorder conditions.

[0058] To provide an overall understanding of the invention, certain illustrative embodiments will now be described. However, it will be understood by one of ordinary skill in the art that the systems and methods described herein can be adapted and modified for other suitable applications and that such other additions and modifications will not depart from the scope hereof.

Definitions

[0059] For convenience, certain terms employed in the specification, examples, and appended claims, are collected here. Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs.

[0060] The articles "a" and "an" are used herein to refer to one or to more than one (i.e., to at least one) of the grammatical object of the article. By way of example, "an element" means one element or more than one element.

[0061] The term "including" is used herein to mean, and is used interchangeably with, the phrase "including but not limited to".

[0062] The term "or" is used herein to mean, and is used interchangeably with, the term "and/or," unless context clearly indicates otherwise.

[0063] The term "such as" is used herein to mean, and is used interchangeably, with the phrase "such as but not limited to".

[0064] A "patient" or "subject" to be treated by the method of the invention can mean either a human or non-human animal, preferably a mammal.

[0065] The term "encoding" comprises an RNA product resulting from transcription of a DNA molecule, a protein resulting from the translation of an RNA molecule, or a protein resulting from the transcription of a DNA molecule and the subsequent translation of the RNA product.

[0066] The term "expression" is used herein to mean the process by which a polypeptide is produced from DNA. The process involves the transcription of the gene into mRNA and the translation of this mRNA into a polypeptide. Depending on the context in which used, "expression" may refer to the production of RNA, protein or both.

[0067] The term "transcriptional regulator" refers to a biochemical element that acts to prevent or inhibit the transcription of a promoter-driven DNA sequence under certain environmental conditions (e.g., a repressor or nuclear inhibitory protein), or to permit or stimulate the transcription of the promoter-driven DNA sequence under certain environmental conditions (e.g., an inducer or an enhancer).

[0068] The terms "array", "microarray", "GeneChip", "genome chip", and "biochip", as used herein refer to an ordered arrangement of hybridizable array elements. The array elements are arranged so that there are preferably at least one or more different array elements on a substrate surface, such as paper, nylon or other type of membrane, filter, chip, glass slide, or any other suitable solid support. The hybridization signal from each of the array elements is individually distinguishable.

[0069] The terms "complementary" or "complementarity" as used herein refer to polynucleotides (i.e., a sequence of nucleotides) related by the base-pairing rules. For example, for the sequence "A-G-T," is complementary to the sequence "T-C-A." Complementarity may be "partial," in which only some of the nucleic acids' bases are matched according to the base pairing rules. Or, there may be "complete" or "total" complementarity between the nucleic acids. The degree of complementarity between nucleic acid strands has significant effects on the efficiency and strength of hybridization between nucleic acid strands. This is of particular importance in amplification reactions, as well as detection methods which depend upon binding between nucleic acids.

[0070] As used herein, the term "hybridization" is used in reference to the pairing of complementary nucleic acids. Hybridization and the strength of hybridization (i.e., the strength of the association between the nucleic acids) is impacted by such factors as the degree of complementarity between the nucleic acids, stringency of the conditions involved, the T_m of the formed hybrid, and the G:C ratio within the nucleic acids.

[0071] As used herein, the term "primer" refers to an oligonucleotide, whether occurring naturally as in a purified restriction digest or produced synthetically, which is capable of acting as a point of

initiation of synthesis when placed under conditions in which synthesis of a primer extension product which is complementary to a nucleic acid strand is induced, (i.e., in the presence of nucleotides and an inducing agent such as DNA polymerase and at a suitable temperature and pH). The primer is preferably single stranded for maximum efficiency in amplification, but may alternatively be double stranded. If double stranded, the primer is first treated to separate its strands before being used to prepare extension products. Preferably, the primer is an oligodeoxy ribonucleotide. The primer must be sufficiently long to prime the synthesis of extension products in the presence of the inducing agent. The exact lengths of the primers will depend on many factors, including temperature, source of primer and the use of the method.

[0072] As used herein, the term "probe" refers to an oligonucleotide (i.e., a sequence of nucleotides), whether occurring naturally as in a purified restriction digest or produced synthetically, recombinantly or by PCR amplification, which is capable of hybridizing to another oligonucleotide of interest. A probe may be single-stranded or double-stranded. Probes are useful in the detection, identification and isolation of particular gene sequences. It is contemplated that any probe used in the present invention will be labeled with any "reporter molecule," so that is detectable in any detection system, including, but not limited to enzyme (e.g., ELISA, as well as enzyme-based histochemical assays), fluorescent, radioactive, and luminescent systems. It is not intended that the present invention be limited to any particular detection system or label.

[0073] As used herein, the terms "compound" and "test compound" refer to any chemical entity, pharmaceutical, drug, and the like that can be used to treat or prevent a disease, illness, conditions, or disorder of bodily function. Compounds comprise both known and potential therapeutic compounds. A compound can be determined to be therapeutic by screening using the screening methods of the present invention. A "known therapeutic compound" refers to a therapeutic compound that has been shown (e.g., through animal trials or prior experience with administration to humans) to be effective in such treatment. In other words, a known therapeutic compound is not limited to a compound efficacious in the treatment of cancer. Examples of test compounds include, but are not limited to peptides, polypeptides, synthetic organic molecules, naturally occurring organic molecules, nucleic acid molecules, and combinations thereof.

[0074] A "sample" may include a single cell or multiple cells or fragments of cells or an aliquot of body fluid or a tissue, taken from a subject, by means including venipuncture, excretion, ejaculation, massage, biopsy, needle aspirate, lavage sample, scraping, surgical incision or intervention or other means known in the art.

[0075] As used herein, the term "subject" refers to an organism, e.g., a human or non-human animal, whether *in vivo*, *ex vivo* or *in vitro*, under observation.

[0076] As used herein, the term "increased expression" refers to the level of a gene expression product that is made higher and/or the activity of the gene expression product that is enhanced. Preferably, the increase is by at least 1.22-fold, 1.5-fold, more preferably the increase is at least 2-fold, 5-fold, or 10-fold, and most preferably, the increase is at least 20-fold, relative to a control.

[0077] As used herein, the term "decreased expression" refers to the level of a gene expression product that is made lower and/or the activity of the gene expression product that is lowered. Preferably, the decrease is at least 25%, more preferably, the decrease is at least 50%, 60%, 70%, 80%, or 90% and most preferably, the decrease is at least one-fold, relative to a control.

[0078] As used herein, the term "gene profile" or "differentially methylated gene profile" refers to an experimentally verified subset of values associated with the methylation level of a set of genes. A differentially methylated gene profile may provide for identification of a biological condition or a physiological process.

[0079] As used herein, the term "differentially methylated gene expression profile," or "gene expression profile" refers to the level or amount of gene expression of particular genes, for example, informative genes, as assessed by methods described herein. The differentially methylated gene expression profile or gene expression profile can comprise data for one or more informative genes and can be measured at a single time point or over a period of time. For example, the differentially methylated gene expression profile or gene expression profile can be determined using a single informative gene, or it can be determined using two or more informative genes, three or more informative genes, five or more informative genes, ten or more informative genes, twenty-five or more informative genes, or fifty or more informative genes. A differentially methylated gene expression profile or gene expression profile may include expression levels of genes that are not informative, as well as informative genes. Phenotype classification (e.g., the presence or absence of a neurological disorder) can be made by comparing the differentially methylated gene expression profile or gene expression profile of the sample with respect to one or more informative genes with one or more differentially methylated gene expression profile or gene expression profiles (e.g., in a database). Using the methods described herein, expression of numerous genes can be measured simultaneously. The assessment of numerous genes provides for a more accurate evaluation of the sample because there are more genes that can assist in classifying the sample. A differentially methylated gene expression profile or gene expression profile may involve only those genes that are increased in expression in a sample, only those genes that are decreased in expression in a sample, or a combination of genes that are increased and decreased in expression in a sample.

[0080] The terms "disorders" and "diseases" are used inclusively and refer to any deviation from the normal structure or function of any part, organ or system of the body (or any combination thereof). A

specific disease is manifested by characteristic symptoms and signs, including biological, chemical and physical changes, and is often associated with a variety of other factors including, but not limited to, demographic, environmental, employment, genetic and medically historical factors. Certain characteristic signs, symptoms, and related factors can be quantitated through a variety of methods to yield important diagnostic information.

[0081] The term "neurological condition" or "neurological disorder" is used herein to mean mental, emotional, or behavioral abnormalities. These include but are not limited to autism spectrum disorder conditions including autism, Asperger's Disorder, bipolar disorder I or II, schizophrenia, schizoaffective disorder, psychosis, depression, stimulant abuse, alcoholism, panic disorder, generalized anxiety disorder, attention deficit disorder, post-traumatic stress disorder, Parkinson's disease, or a combination thereof.

Gene Chips

[0082] One aspect of the invention provides gene chips. Gene chips, also called "biochips" or "arrays" or "microarrays" are miniaturized devices typically with dimensions in the micrometer to millimeter range for performing chemical and biochemical reactions and are particularly suited for embodiments of the invention. Arrays may be constructed via microelectronic and/or microfabrication using essentially any and all techniques known and available in the semiconductor industry and/or in the biochemistry industry, provided that such techniques are amenable to and compatible with the deposition and screening of polynucleotide sequences. Microarrays are particularly desirable for their virtues of high sample throughput and low cost for generating profiles and other data.

[0083] One specific aspect of the invention provides a gene chip having a plurality of different oligonucleotides having specificity for DNA-methylated genes or fragments thereof associated with neurological conditions, and in particular, autism spectrum disorder conditions including pervasive developmental disorder-not otherwise specified (PDD-NOS), including atypical autism, Asperger's Disorder, or a combination thereof. In a related embodiment, the invention provides a gene chip having a plurality of different oligonucleotides having specificity for genes whose DNA methylation level changes in a subject who is afflicted with neurological conditions, and in particular, autism spectrum disorder conditions including pervasive developmental disorder-not otherwise specified (PDD-NOS), including atypical autism, Asperger's Disorder, or a combination thereof when the subject responds favorably to a therapeutic treatment that is intended to treat the neurological condition.

[0084] In one embodiment of the gene chips provided herein, the oligonucleotides on the gene chip comprise a plurality oligonucleotides that are specific for the differentially DNA-methylated genes, or promoter regions or CpG islands thereof, set out in Tables 1, 2, 3, or combinations thereof. In another

embodiment, the gene chip has oligonucleotides specific for the genes associated with autism spectrum disorder conditions including pervasive developmental disorder-not otherwise specified (PDD-NOS), including atypical autism, Asperger's Disorder, or a combination thereof.

[0085] In another specific embodiment, the gene chip has at least one oligonucleotide specific for genes associated with the cellular response to androgens.

[0086] In another specific embodiment, the gene chip has at least one oligonucleotide specific for genes associated with circadian rhythm. In another specific embodiment, the gene chip has at least one oligonucleotide specific for the circadian rhythm associated genes, or any of the genes set out in Table 1, Table 2, or Table 3, or any combination thereof.

[0087] In another specific embodiment, the gene chip has at least one oligonucleotide specific for target genes associated with nervous system development, WNT signaling, axon guidance, regulation of the cytoskeleton, synaptic transmission or plasticity, myelination, long-term potentiation, neuron toxicity, Purkinje cell differentiation, cerebellar development, embryonic development, regulation of actin networks, digestion, inflammation, oxidative stress, epilepsy, apoptosis, morphogenesis, cell survival, differentiation, the unfolded protein response, Type II diabetes and insulin signaling, digestion, liver toxicity (hepatic stellate cell activation, fibrosis, and cholestasis), epilepsy, apoptosis, cell survival, differentiation, the unfolded protein response, endocrine function, circadian rhythm, cholesterol metabolism and the steroidogenesis pathway, or a combination thereof.

[0088] In another embodiment, the gene chip comprises oligonucleotide probes specific for genes associated with apoptosis and inflammation, as well as many neurological and metabolic processes commonly associated with ASD, such as myelination, neuron plasticity, synaptic transmission, and hypercholesterolemia. In one embodiment, the gene chip comprises oligonucleotides specific for BCL-2, RORA, or a combination thereof.

[0089] In another specific embodiment of the gene chips provided herein, the gene chip comprises at least 3, 5, 10, 15, 20 or 25 of the probes are derived from oligonucleotides that are specific for the differentially DNA-methylated genes set out in any one of Tables 1, 2, 3, or a combination thereof. In a related embodiment, at least 50% of the probes on the gene chip are derived from oligonucleotides that are specific for the differentially DNA-methylated genes present in any one of Tables 1, 2, 3, or a combination thereof. In a related embodiment, at least 70%, 80%, 90%, 95% or 98% of the probes on the gene chip are derived from oligonucleotides that are specific for the differentially DNA-methylated genes present in any one of Tables 1, 2, 3, or combinations thereof.

[0090] The invention further provides a gene chip for distinguishing cell samples from individuals

having a positive prognosis and cell samples from individuals having a negative prognosis, wherein prognosis refers to the progression of disease or prognosis for successful treatment by a given treatment regimen or agent, comprising a positionally-addressable array of polynucleotide probes bound to a support, said polynucleotide probes comprising a plurality of polynucleotide probes of different nucleotide sequences, each of said different nucleotide sequences comprising a sequence complementary and specifically hybridizable to at least one of a plurality of differentially DNA-methylated genes, said plurality consisting of at least one of the differentially DNA-methylated genes listed in Tables 1, 2, 3, or a combination thereof.

[0091] In some embodiments of the gene chips, processes, methods and kits provided by the invention, the neurological condition is selected from the group consisting of autism spectrum disorders, autism, atypical autism, pervasive developmental disorder-not otherwise specified (PDD-NOS), Asperger's Disorder, Rett's syndrome, allodynia, catalepsy, hypernocieption, Parkinson's disease, parkinsonism, cognitive impairments, age-associated memory impairments, cognitive impairments, dementia associated with neurologic and/or neurological conditions, allodynia, catalepsy, hypernocieption, and epilepsy, brain tumors, brain lesions, multiple sclerosis, Down's syndrome, progressive supranuclear palsy, frontal lobe syndrome, schizophrenia, delirium, Tourette's syndrome, myasthenia gravis, attention deficit hyperactivity disorder, dyslexia, mania, depression, apathy, myopathy, Alzheimer's disease, Huntington's Disease, dementia, encephalopathy, schizophrenia, severe clinical depression, brain injury, Attention Deficit Disorder (ADD), Attention Deficit Hyperactivity Disorder (ADHD), hyperactivity disorder, Asperger's Disorder, bipolar manic-depressive disorder, ischemia, alcohol addiction, drug addiction, obsessive compulsive disorders, Pick's disease and Binswanger's disease.

[0092] DNA microarray and methods of analyzing data from microarrays are well-described in the art, including in DNA Microarrays: A Molecular Cloning Manual, Ed by Bowtell and Sambrook (Cold Spring Harbor Laboratory Press, 2002); Microarrays for an Integrative Genomics by Kohana (MIT Press, 2002); A Biologist's Guide to Analysis of DNA Microarray Data, by Knudsen (Wiley, John & Sons, Incorporated, 2002); and DNA Microarrays: A Practical Approach, Vol. 205 by Schema (Oxford University Press, 1999); and Methods of Microarray Data Analysis II, ed. by Lin et al. (Kluwer Academic Publishers, 2002), hereby incorporated by reference in their entirety.

[0093] Microarrays may be prepared by selecting probes which comprise a polynucleotide sequence, and then immobilizing such probes to a solid support or surface. For example, the probes may comprise DNA sequences, RNA sequences, or copolymer sequences of DNA and RNA. The polynucleotide sequences of the probes may also comprise DNA and/or RNA analogues, or combinations thereof. For example, the polynucleotide sequences of the probes may be full or partial fragments of genomic DNA.

The polynucleotide sequences of the probes may also be synthesized nucleotide sequences, such as synthetic oligonucleotide sequences. The probe sequences can be synthesized either enzymatically *in vivo*, enzymatically *in vitro* (e.g., by PCR), or non-enzymatically *in vitro*.

[0094] The probe or probes used in the methods and gene chips of the invention may be immobilized to a solid support which may be either porous or non-porous. For example, the probes of the invention may be polynucleotide sequences which are attached to a nitrocellulose or nylon membrane or filter covalently at either the 3' or the 5' end of the polynucleotide. Such hybridization probes are well known in the art (see, e.g., Sambrook et al., MOLECULAR CLONING--A LABORATORY MANUAL (2ND ED.), Vols. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y. (1989). Alternatively, the solid support or surface may be a glass or plastic surface. In one embodiment, hybridization levels are measured to microarrays of probes consisting of a solid phase on the surface of which are immobilized a population of polynucleotides, such as a population of DNA or DNA mimics, or, alternatively, a population of RNA or RNA mimics. The solid phase may be a nonporous or, optionally, a porous material such as a gel.

[0095] In one embodiment, a microarray comprises a support or surface with an ordered array of binding (e.g., hybridization) sites or "probes" each representing one of the markers described herein. Preferably the microarrays are addressable arrays, and more preferably positionally addressable arrays. More specifically, each probe of the array is preferably located at a known, predetermined position on the solid support such that the identity (i.e., the sequence) of each probe can be determined from its position in the array (i.e., on the support or surface). In preferred embodiments, each probe is covalently attached to the solid support at a single site.

[0096] Microarrays can be made in a number of ways, of which several are described below. However produced, microarrays share certain characteristics. The arrays are reproducible, allowing multiple copies of a given array to be produced and easily compared with each other. Preferably, microarrays are made from materials that are stable under binding (e.g., nucleic acid hybridization) conditions. The microarrays are preferably small, e.g., between 1 cm² and 25 cm², between 12 cm² and 13 cm², or about 3 cm². However, larger arrays are also contemplated and may be preferable, e.g., for use in screening arrays. Preferably, a given binding site or unique set of binding sites in the microarray will specifically bind (e.g., hybridize) to the product of a single gene in a cell (e.g., to a specific mRNA). However, in general, other related or similar sequences will cross hybridize to a given binding site.

[0097] The microarrays of the present invention include one or more test probes, each of which has a polynucleotide sequence that is complementary to a subsequence of RNA or DNA to be detected. Preferably, the position of each probe on the solid surface is known. Indeed, the microarrays are preferably positionally addressable arrays. Specifically, each probe of the array is preferably located at

a known, predetermined position on the solid support such that the identity (i.e., the sequence) of each probe can be determined from its position on the array (i.e., on the support or surface).

[0098] According to one aspect of the invention, the microarray is an array (i.e., a matrix) in which each position represents one of the markers or gene biomarkers as described herein. For example, each position can contain a DNA or DNA analogue based on genomic DNA to which a particular RNA transcribed from that genetic marker or biomarker can specifically hybridize. The DNA or DNA analogue can be, for example, a synthetic oligomer or a gene fragment. In one embodiment, probes representing each of the differentially DNA-methylated genes or biomarkers in Tables 1, 2, 3, or a combination thereof are present on the array.

[0099] As noted above, the "probe" to which a particular polynucleotide molecule specifically hybridizes according to the invention contains a complementary polynucleotide sequence. In one embodiment, the probes of the differentially DNA-methylated gene array consist of nucleotide sequences of 10 to 1,000 nucleotides. In a preferred embodiment, the nucleotide sequences of the probes are in the range of 10-200 nucleotides in length and are genomic sequences of a species of organism, such that a plurality of different probes is present, with sequences complementary and thus capable of hybridizing to the genome of such a species of organism, sequentially tiled across all or a portion of such genome. In other specific embodiments, the probes are in the range of 10-30 nucleotides in length, in the range of 10-40 nucleotides in length, in the range of 20-50 nucleotides in length, in the range of 40-80 nucleotides in length, in the range of 50-150 nucleotides in length, in the range of 80-120 nucleotides in length, and most preferably are 60 nucleotides in length.

[0100] The probes may comprise DNA or DNA "mimics" (e.g., derivatives and analogues) corresponding to a portion of an organism's genome. In another embodiment, the probes of the microarray are complementary RNA or RNA mimics. DNA mimics are polymers composed of subunits capable of specific, Watson-Crick-like hybridization with DNA, or of specific hybridization with RNA. The nucleic acids can be modified at the base moiety, at the sugar moiety, or at the phosphate backbone. Exemplary DNA mimics include, e.g., phosphorothioates.

[0101] DNA can be obtained, e.g., by polymerase chain reaction (PCR) amplification of genomic DNA or cloned sequences. PCR primers are preferably chosen based on a known sequence of the genome that will result in amplification of specific fragments of genomic DNA. Computer programs that are well known in the art are useful in the design of primers with the required specificity and optimal amplification properties, such as OLIGO version 5.0 (National Biosciences, Inc. Plymouth, MN). Typically each probe on the microarray will be between 10 bases and 50,000 bases, usually

between 300 bases and 1,000 bases in length. PCR methods are well known in the art, and are described, for example, in Innis et al., eds., PCR: Protocols: A Guide to Methods and Applications, Academic Press Inc., San Diego, Calif. (1990). It will be apparent to one skilled in the art that controlled robotic systems are useful for isolating and amplifying nucleic acids.

[0102] An alternative, means for generating the polynucleotide probes of the microarray is by synthesis of synthetic polynucleotides or oligonucleotides, e.g., using N-phosphonate or phosphoramidite chemistries (Froehler et al., Nucleic Acid Res. 14:5399- 5407 (1986); McBride et al., Tetrahedron Lett. 24:246-248 (1983)). Synthetic sequences are typically between about 10 and about 500 bases in length, more typically between about 20 and about 100 bases, and most preferably between about 40 and about 70 bases in length. In some embodiments, synthetic nucleic acids include non-natural bases, such as, but by no means limited to, inosine. As noted above, nucleic acid analogues may be used as binding sites for hybridization. An example of a suitable nucleic acid analogue is peptide nucleic acid and methods for the preparation of peptide nucleic acids are known in the art (see, e.g., Egholm et al., Nature 363:566-568 (1993); U.S. Pat. No. 5,539,083). Probes are preferably selected using an algorithm that takes into account binding energies, base composition, sequence complexity, cross-hybridization binding energies, and secondary structure (see Friend et al., International Patent Publication WO 01/05935, published Jan. 25, 2001; Hughes et al., Nat. Biotech. 19:342-7 (2001)).

[0103] A skilled artisan will also appreciate that positive control probes, e.g., probes known to be complementary and hybridizable to sequences in the DNA molecules, and negative control probes, e.g., probes known to not be complementary and hybridizable to sequences in the DNA molecules, should be included on the array. In one embodiment, positive controls are synthesized along the perimeter of the array. In another embodiment, positive controls are synthesized in diagonal stripes across the array. In still another embodiment, the reverse complement for each probe is synthesized next to the position of the probe to serve as a negative control. In yet another embodiment, sequences from other species of organism are used as negative controls or as "spike-in" controls.

[0104] The probes may be attached to a solid support or surface, which may be made, e.g., from glass, plastic (e.g., polypropylene, nylon), polyacrylamide, nitrocellulose, gel, or other porous or nonporous material. A preferred method for attaching the nucleic acids to a surface is by printing on glass plates, as is described generally by Schena et al., Science 270:467-470 (1995). This method is especially useful for preparing microarrays of cDNA (See also, DeRisi et al., Nature Genetics 14:457-460 (1996); Shalon et al., Genome Res. 6:639-645 (1996); and Schena et al., Proc. Natl. Acad. Sci. U.S.A. 93:10539-11286 (1995)).

[0105] A second preferred method for making microarrays is by making high-density oligonucleotide

arrays. Techniques are known for producing arrays containing thousands of oligonucleotides complementary to defined sequences, at defined locations on a surface using photolithographic techniques for synthesis in situ (see, Fodoret al., 1991, Science 251:767-773; Pease et al., 1994, Proc. Natl. Acad. Sci. U.S.A. 91:5022-5026; Lockhart et al., 1996, Nature Biotechnology 14:1675; U.S. Pat. Nos. 5,578,832; 5,556,752; and 5,510,270) or other methods for rapid synthesis and deposition of defined oligonucleotides (Blanchard et al., Biosensors & Bioelectronics 11:687-690). When these methods are used, oligonucleotides (e.g., 60-mers) of known sequence are synthesized directly on a surface such as a derivatized glass slide. Usually, the array produced is redundant, with several oligonucleotide molecules per RNA.

[0106] Other methods for making microarrays, e.g., by masking (Maskos and Southern, 1992, Nuc. Acids. Res. 20:1679-1684), may also be used. In principle, and as noted supra, any type of array, for example, dot blots on a nylon hybridization membrane (see Sambrook et al., MOLECULAR CLONING--A LABORATORY MANUAL (2ND ED.), Vols. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y. (1989)) could be used. However, as will be recognized by those skilled in the art, very small arrays will frequently be preferred because hybridization volumes will be smaller. In one embodiment, the arrays of the present invention are prepared by synthesizing polynucleotide probes on a support. In such an embodiment, polynucleotide probes are attached to the support covalently at either the 3' or the 5' end of the polynucleotide.

[0107] In a one embodiment, microarrays of the invention are manufactured by means of an ink jet printing device for oligonucleotide synthesis, e.g., using the methods and systems described by Blanchard in U.S. Pat. No. 6,028,189; Blanchard et al., 1996, Biosensors and Bioelectronics 11:687-690; Blanchard, 1998, in SYNTHETIC DNA ARRAYS IN GENETIC ENGINEERING, Vol. 20, J. K. Setlow, Ed., Plenum Press, New York at pages 111-123. Specifically, the oligonucleotide probes in such microarrays are preferably synthesized in arrays, e.g., on a glass slide, by serially depositing individual nucleotide bases in "microdroplets" of a high surface tension solvent such as propylene carbonate. The microdroplets have small volumes (e.g., 100 pL or less, more preferably 50 pL or less) and are separated from each other on the microarray (e.g., by hydrophobic domains) to form circular surface tension wells which define the locations of the array elements (i.e., the different probes). Microarrays manufactured by this ink-jet method are typically of high density, preferably having a density of at least about 2,500 different probes per 1 cm². The polynucleotide probes are attached to the support covalently at either the 3' or the 5' end of the polynucleotide.

[0108] Also an embodiment of this invention are methods of screening a subject for a neurological disease or disorder, and assessing a subject's risk for having or developing a neurological disease or disorder, e.g., and autism spectrum disorder. The methods comprise isolating nucleic acid, e.g., a DNA sample,

from a sample from a test subject, and assaying the differential DNA methylation of at least one gene in Table 1, Table 2, Table 3 or a combination thereof in the isolated DNA of a subject and then comparing the level of methylation of the genes in the test sample to the methylation of those same genes in a control sample. Preferably the gene(s) comprise RORA or RORA and BCL-2. Where the methylation profile from the test subject is compared to a negative control, a statistically significant difference in the differential methylation of at least one gene listed in Table 1, Table 2, Table 3 between the test sample and the negative control aides in assessing the test subject's likelihood of having or developing the neurological disease or disorder. For example, a difference of log₂ ratio greater than ± 0.3 between the level of methylation of the gene(s) of the test sample and the same genes in the control sample indicates an increased risk for having or developing a neurological disorder or disease, e.g. an autism spectrum disorder. The a negative control may be, e.g., a nucleic acid, DNA, sample from a plurality of subjects who do not have a neurological disease or disorder, particularly the neurological disorder for which the subject is being tested, e.g., an autism spectrum disorder. The control may be, e.g., non-phenotypic discordant twin, siblings of the subject, or unrelated subjects. The autism spectrum disorder may comprise, or be selected from the group consisting of, pervasive developmental disorder-not otherwise specified (PDD-NOS), including atypical autism, and Asperger's Disorder. The autism spectrum disorder may also be characterized by a lower severity score across all ADIR items, or an intermediate severity score across all ADIR items, or a higher severity scores on spoken language items on the ADIR, or a higher frequency of savant skills, or a severe language impairment, or a combination thereof.

[0109] The nucleic acid may be obtained from a neurological sample as described herein. The nucleic acid may be obtained from a peripheral tissue sample, e.g., epithelial cells from a cheek swab, or a lymphocyte containing sample, e.g., blood. The nucleic acid may also be obtained from a lymphoblastoid cell line (LCL) derived from the test or control subject.

[0110] The level of DNA methylation in the nucleic acid samples may be quantified by methods known in the art some of which are disclosed herein, e.g., bisulfite-sequencing, methylation-specific PCR (MSP), large scale CpG island microarray analysis using methylation-enriched and -unenriched samples, promoter analysis using methylation-enriched and -unenriched samples, combined bisulfite and restriction analysis (COBRA), pyrosequencing, Methyl-Light (Premier Biosoft International, Palo Alto CA) or Methyl-Profiler (SABiosciences, Frederick, MD) methylation analysis, Sequenom mass analyzer, bisulfite treatment of DNA combined with selected amplification of the specific promoter regions of differentially methylated genes with the addition of a T7 promoter tag, followed by T7-mediated transcription and RNase T1 cleavage of the transcript with methylation sites determined by MALDI-TOF mass spectrometry analysis in the manner of Schatz et al. (2004), or a combination thereof.

Methods of Determining Differentially DNA-Methylated Gene Profiles

[0111] One aspect of the invention provides methods for determining a differentially DNA-methylated gene profile for a specific neurological disorder or neurological condition, such as autism spectrum disorder conditions including autistic disorder, pervasive developmental disorder-not otherwise specified (PDD-NOS), including atypical autism, Asperger's Disorder. Furthermore, the systems and methods described herein may be employed to generate differentially DNA-methylated gene profiles for diseases or disorders of interest. This expression data may be analyzed independently to determine a differentially DNA-methylated gene profile of interest, or combined with the existing biological data stored in a plurality of different types of databases. Statistical analyses may be applied as well as machine learning techniques that are used to discover trends and patterns in the underlying data. These techniques include clustering methods, which can be used for example to organize microarray expression data.

[0112] One specific aspect of the invention provides a method for determining a differentially DNA-methylated gene profile for a neurological condition, comprising (i) preparing samples of control and experimental DNA, wherein the experimental DNA is generated from a nucleic acid sample isolated from a subject suspected of being afflicted with the neurological condition; (ii) preparing one or more microarrays comprising a plurality of different oligonucleotides having specificity for differentially DNA-methylated genes associated with the neurological condition; (iii) applying the prepared samples to the one or more microarrays to allow hybridization between the oligonucleotides and the control and experimental DNA s; (v) identifying the oligonucleotides on the microarray which display differential hybridization to the experimental DNA relative to the control DNA; and (vi) identifying a set of genes from the oligonucleotides identified in step (v) thereby determining a differentially DNA-methylated gene profile for the neurological condition.

[0113] In a preferred embodiment, the neurological condition is an autism spectrum disorder condition including autistic disorder, pervasive developmental disorder-not otherwise specified (PDD-NOS), including atypical autism, Asperger's Disorder, or a combination thereof. In another embodiment, the neurological condition is selected from the group consisting of autism spectrum disorder conditions including autistic disorder, pervasive developmental disorder-not otherwise specified (PDD-NOS), including atypical autism, Asperger's Disorder, Rett's syndrome, Parkinson's disease, parkinsonism, cognitive impairments, age-associated memory impairments, cognitive impairments, dementia associated with neurologic and/or neurological conditions, allodynia, catalepsy, hypernocieption, and epilepsy, brain tumors, brain lesions, multiple sclerosis, Down's syndrome, progressive supranuclear palsy, frontal lobe syndrome, schizophrenia, delirium, Tourette's syndrome, myasthenia gravis, attention deficit hyperactivity disorder, dyslexia, mania, depression, apathy, myopathy, Alzheimer's disease, Huntington's Disease, dementia, encephalopathy, schizophrenia, severe clinical depression, brain injury, Attention Deficit Disorder (ADD), Attention Deficit Hyperactivity Disorder (ADHD),

hyperactivity disorder, bipolar manic-depressive disorder, ischemia, alcohol addiction, drug addiction, obsessive compulsive disorders, Pick's disease and Binswanger's disease.

[0114] In another embodiment, the samples of experimental DNA may be isolated from a subject or group of subjects suspected of being afflicted or afflicted with one or more neurological conditions, such as autism spectrum disorder condition including autistic disorder, pervasive developmental disorder-not otherwise specified (PDD-NOS), including atypical autism, Asperger's Disorder, or a combination thereof. Control DNA may be derived from a nucleic acid sample of a subject or group of subjects which are not afflicted with the neurological conditions that the subjects from which the experimental DNA was derived. In another embodiment, the subjects from which the experimental and control samples are derived may both be suspected of being afflicted or afflicted with the condition, but the severity of the condition or a treatment plan in the two subject groups may differ.

[0115] A related aspect of the invention provides a method of determining a differentially DNA-methylated gene profile for the administration of a therapeutic treatment to a subject. Such methods are useful to detect the gene expression changes that accompany the underlying therapeutic treatments. A differentially DNA-methylated gene profile for such genetic changes may be used to determine if a second therapeutic treatment is expected to have the same effect, by comparing the differentially DNA-methylated gene profile of the second treatment to the differentially DNA-methylated gene profile of the first.

[0116] Accordingly, one specific aspect of the invention provides a method of determining a differentially DNA-methylated gene profile indicative for the administration of a therapeutic treatment to a subject, the method comprising (i) preparing samples of control and experimental DNA, wherein the experimental DNA is generated from a nucleic acid sample isolated from a subject who has received or is receiving the therapeutic treatment; (ii) preparing one or more microarrays comprising a plurality of different oligonucleotides wherein the oligonucleotides are specific to differentially DNA-methylated genes associated with an autism spectrum disorder; (iii) applying the prepared samples to the one or more microarrays to allow hybridization between the oligonucleotides and the control and experimental DNAs; (v) identifying the oligonucleotides on the microarray which display differential hybridization to the experimental DNA relative to the control DNA; (vi) identifying a set of genes or differentially DNA-methylated genes associated with an autism spectrum disorder from the oligonucleotides identified in step (v) thereby determining a differentially DNA-methylated gene profile for the administration of the therapeutic treatment to the subject.

[0117] In yet another aspect of the invention, a method is provided for determining a differentially DNA-methylated gene profile for at least one autism spectrum disorder, comprising (a) preparing samples of control and experimental DNA, wherein the experimental DNA is generated from a nucleic

acid sample isolated from a subject suspected of being afflicted with the at least one autism spectrum disorder and the control DNA is generated from a nucleic acid sample isolated from a healthy individual; (b) preparing one or more microarrays comprising a plurality of different oligonucleotides having specificity for differentially DNA-methylated genes associated with the at least one autism spectrum disorder; (c) applying the prepared samples to the one or more microarrays to allow hybridization between the oligonucleotides and the control DNA s and the oligonucleotide and the experimental DNA s; (d) identifying the oligonucleotides on the microarray which display differential hybridization to the experimental DNA relative to the control DNA thereby determining a differentially DNA-methylated gene profile for the at least one autism spectrum disorder.

[0118] In yet another aspect of the invention, a method is provided for distinguishing between different phenotypes of an autism spectrum disorder comprising severely language impaired (L), mildly affected (M), or “savants” (S) comprising (a) preparing samples of control and experimental DNA, wherein the experimental DNA is generated from a nucleic acid sample isolated from a subject suspected of being afflicted with at least one phenotype comprising the severely language impaired (L), mildly affected (M), or “savants” (S); (b) preparing one or more microarrays comprising a plurality of different oligonucleotides having specificity for differentially DNA-methylated genes associated with the at least one phenotype; (c) applying the prepared samples to the one or more microarrays to allow hybridization between the oligonucleotides and the control and experimental DNA s; (d) identifying the oligonucleotides on the microarray which display differential hybridization to the experimental DNA relative to the control DNA thereby determining a differentially DNA-methylated gene profile for distinguishing among the different phenotypes of autism spectrum disorder. The control may be e.g., a sample from a subject who is known to have an autism spectrum disorder, e.g., severely language impaired (L), mildly affected (M), or “savants” (S), wherein a greater similarity of the experimental profile to one of the ASD profiles more than to the others distinguishes the phenotype of the experimental profile.

Methods for categorizing subjects as severely language impaired (L), mildly affected (M), or “savants” (S)

[0119] In yet another embodiment of the screening method of the present invention, the method distinguishes between different variants of autism spectrum disorder comprising a lower severity scores across all ADIR items, an intermediate severity across all ADIR items, a higher severity scores on spoken language items on the ADIR, a higher frequency of savant skills, and a severe language impairment, or a combination thereof. Methods for determining severity scores across ADIR items, savant skills and language impairment are known. See, e.g., Hu and Steinberg (2009) Autism Research 2:67-77, e.g., Tables 1 and Supplementary Table 1 therein, wherein ADIR scores of about 2000 autistic individuals were used to identify phenotypic subgroups with idiopathic ASD who were

characterized by combined symptoms across multiple domains.

[0120] In one embodiment of the methods for determining a differentially DNA-methylated gene profile for the administration of a therapeutic treatment, administration of therapeutic treatment results in a physiological change in the subject, such as a beneficial change. In a specific embodiment, the physiological change comprises one or more improvements in social interaction, language abilities, restricted interests, repetitive behaviors, sleep disorders, seizures, gastrointestinal, hepatic, and mitochondrial function, neural inflammation, or a combination thereof. In another embodiment, the control DNA may be derived from the subject(s) prior to administration of the therapeutic treatment, or from a subject or group of subjects who do not receive the therapeutic treatment.

[0121] In another embodiment of the methods for determining a differentially DNA-methylated gene profile for the administration of a therapeutic treatment to a subject suspected of being afflicted with or afflicted with autism spectrum disorder conditions including autistic disorder, pervasive developmental disorder-not otherwise specified (PDD-NOS), including atypical autism, Asperger's Disorder, the therapeutic treatment may comprise a single procedure or it may comprise an aggregate of treatment procedures. In one embodiment, therapeutic treatment comprises a behavioral therapy, such as applied behavior analysis (ABA) intervention methods, dietary changes, exercise, massage therapy, group therapy, talk therapy, play therapy, conditioning, or alternative therapies such as sensory integration and auditory integration therapies. In another embodiment, the therapeutic treatment comprises administering to the subject a drug, such as an antidepressant or antipsychotic drug. In another embodiment, the subject is afflicted with a neurological condition other than autism spectrum disorder conditions including autistic disorder, pervasive developmental disorder-not otherwise specified (PDD-NOS), including atypical autism, Asperger's Disorder. Such condition may be one which the therapeutic treatment is intended to treat. In another embodiment, the subject is a healthy subject who is not afflicted with a neurological condition. In another embodiment, the therapeutic treatment is a treatment for the autism spectrum disorder neurological conditions including autistic disorder, pervasive developmental disorder-not otherwise specified (PDD-NOS), including atypical autism, Asperger's Disorder.

[0122] In another embodiment, the drug being administered in the single procedure or the aggregate of treatment procedures is a serotonergic antidepressant medication, such as one selected from the group consisting of citalopram, fluoxetine, fluvoxamine, paroxetine, or sertraline, or the drug is a catecholaminergic antidepressant medication, such as bupropion.

[0123] In another preferred embodiment of the methods described herein, both the control DNA and the experimental DNA are derived from a nucleic acid sample isolated from the subject. Samples may be isolated from a mammal, such as a human. In a specific embodiment, the sample is isolated post-

mortem from a human. Nucleic acid samples may be isolated from any tissue or bodily fluid, including blood, saliva, tears, cerebrospinal fluid, pericardial fluid, synovial fluid, amniotic fluid, semen, bile, ear wax, gastric acid, sweat, urine, or fluid drained from an edema. In a further specific embodiment, the nucleic acid sample is isolated from lymphoblastoid cells or lymphoblastoid cell lines (LCL) derived from blood cells of subjects. In some embodiments of the methods described herein, the sample is isolated from a neuronal tissue or a combination of tissue types, such as olfactory bulb cells, cerebrospinal fluid, hypothalamus, amygdala, pituitary, spinal cord, brainstem, cerebellum, cortex, frontal cortex, hippocampus, choroid plexus, striatum, and thalamus.

[0124] In one embodiment of the methods described herein, the microarray is any one of the microarrays, or differentially DNA-methylated gene chips described herein. In a preferred embodiment, the oligonucleotides on the microarray comprise those specific to differentially DNA-methylated genes selected from Table 1, Table 2, Table 3, or a combination thereof. In a specific embodiment, the oligonucleotides of the microarray are specific to differentially DNA-methylated genes associated with nervous system development, axon guidance, synaptic transmission or plasticity, myelination, long-term potentiation, neuron toxicity, Purkinje cell differentiation, cerebellar development, embryonic development, regulation of actin networks, digestion, inflammation, oxidative stress, epilepsy, apoptosis, morphogenesis, cell survival, differentiation, the unfolded protein response, Type II diabetes and insulin signaling, digestion, liver toxicity (hepatic stellate cell activation, fibrosis, and cholestasis), endocrine function, circadian rhythm, cholesterol metabolism and the steroidogenesis pathway, or a combination thereof as described supra. In a preferred embodiment, at least 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, or 99% of the genes on the microarray are specific to differentially DNA-methylated genes selected from Table 1, Table 2, Table 3, or a combination thereof.

[0125] In another embodiment of the methods described herein, the control DNA and the experimental DNA are hybridized to the same microarray, while in another embodiment they are hybridized to separate but substantially identical microarrays. If the same microarray is used, the DNA samples may be labeled using fluorescent compounds having different emission wavelengths such that the signals generated by each DNA type may be distinguished from a single microarray.

[0126] In yet another embodiment of the methods described herein, the control and experimental DNA is isolated from one or more subjects. In one embodiment, the control DNA and experimental DNA are isolated each from at least 3, 5, 10, 15 or 20 subjects. The DNAs from each subject may be hybridized to the microarrays separately, or the control DNAs, or the experimental DNAs, may be pooled together, such that, for example, an experimental DNA sample is derived from multiple subjects. In preferred embodiments, the subjects are mammals, such as rodents, primates or humans.

[0127] In one embodiment of the methods described herein, the set of differentially DNA-methylated genes in the gene profile comprise differentially DNA-methylated genes which have a differential expression in the experimental DNA relative to the control DNA. Differential expression may refer to a lower expression level or to a higher expression. In preferred embodiments, the difference in expression level is statistically significant for each differentially DNA-methylated gene, or marker, on the set. In preferred embodiments, the difference in expression is at least 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%, 150%, 200%, 300%, 400%, or 500% greater in the experimental DNA than in the control DNA, or vice versa. In another preferred embodiment, the difference in expression is at least about 1.22-fold, 1.5-fold, 2-fold, 2.5-fold, 3-fold, 3.5-fold, 4-fold, 4.5-fold, 5-fold, 6-fold, 7-fold, 8-fold, 9-fold, 10-fold, 12-fold, 14-fold, 16-fold, 18-fold, 20-fold, 25-fold, 30-fold, 35-fold, 40-fold, 45-fold, 50-fold, 55-fold, 60-fold, 65-fold, 70-fold, 75-fold, 80-fold, 85-fold, 90-fold, 95-fold, 100-fold greater (or intermediate ranges thereof as another example) in the experimental DNA than in the control DNA, or vice versa. A gene profile may comprise all the differentially DNA-methylated genes which are differentially expressed between the control and experimental DNAs or it may comprise a subset of those genes. In some embodiments, the gene profile comprises at least 1%, 2%, 3%, 4%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 98%, 99% or 100% (or intermediate ranges thereof as another example) of the differentially DNA-methylated genes having differential expression. Differentially DNA-methylated genes showing large, reproducible changes in expression between the two samples are preferred in some embodiments. In preferred embodiments, the differentially DNA-methylated gene profile further comprises a subset of values associated with the expression level of each of the differentially DNA-methylated gene in the profile, such that differentially DNA-methylated gene profile allows the identification of a biological and/or pathological condition, an agent and/or its biological mechanism of action, or a physiological process.

[0128] The preparation of samples of control and experimental DNA may be carried out using techniques known in the art. The DNA molecules analyzed by the present invention may be from any clinically relevant source. In one embodiment, the DNA is derived from RNA, including, but by no means limited to, total cellular RNA, poly(A)⁺ messenger RNA (mRNA) or fraction thereof, cytoplasmic mRNA, or RNA transcribed from cDNA (i.e., cRNA; see, e.g., U.S. Pat. Nos. 5,545,522, 5,891,636, or 5,716,785). Methods for preparing total and poly(A)⁺ RNA are well known in the art, and are described generally, e.g., in Sambrook et al., *MOLECULAR CLONING--A LABORATORY MANUAL* (2ND ED.), Vols. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y. (1989). In one embodiment, RNA is extracted from a sample of cells of the various tissue types of interest, such as the lymphoblastoid cell or lymphoblastoid cell line derived therefrom or from the aforementioned neuronal tissue types, using guanidinium thiocyanate lysis followed by CsCl centrifugation (Chirgwin et al., 1979, *Biochemistry* 18:5294-5299). In another embodiment, total RNA is extracted using a silica gel-based column, commercially available examples of which include

RNeasyTM (Qiagen, Valencia, Calif.) and StrataPrep (Stratagene, La Jolla, Calif.). Poly(A)⁺ RNA can be selected, e.g., by selection with oligo-dT cellulose or, alternatively, by oligo-dT primed reverse transcription of total cellular RNA. In one embodiment, RNA can be fragmented by methods known in the art, e.g., by incubation with ZnCl₂, to generate fragments of RNA. In another embodiment, the polynucleotide molecules analyzed by the invention comprise PCR products of amplified polynucleotides (e.g., RNA or cDNA, among others). DNA molecules that are poorly expressed in particular cells may be enriched using normalization techniques (Bonaldo et al., 1996, Genome Res. 6:791-806).

[0129] The DNAs may be detectably labeled at one or more nucleotides. Any method known in the art may be used to detectably label the DNAs. Preferably, this labeling incorporates the label uniformly along the length of the RNA, and more preferably, the labeling is carried out at a high degree of efficiency. One embodiment for this labeling uses oligo-dT primed reverse transcription to incorporate the label; however, conventional methods of this method are biased toward generating 3' end fragments. Thus, in one embodiment, random primers (e.g., 9-mers) are used in reverse transcription to uniformly incorporate labeled nucleotides over the full length of the DNAs. Alternatively, random primers may be used in conjunction with PCR methods or T7 promoter-based in vitro transcription methods in order to amplify the cDNAs.

[0130] In one embodiment, the detectable label is a luminescent label. For example, fluorescent labels, bioluminescent labels, chemiluminescent labels, and colorimetric labels may be used in the present invention. In one preferred embodiment, the label is a fluorescent label, such as a fluorescein, a phosphor, a rhodamine, or a polymethine dye derivative. Examples of commercially available fluorescent labels include, for example, fluorescent phosphoramidites such as FLUOREPRIMETM (Amersham Pharmacia, Piscataway, N.J.), Fluoredate (Millipore, Bedford, Mass.), FAM (ABI, Foster City, Calif.), and Cy3 or Cy5 (Amersham Pharmacia, Piscataway, N.J.). In another embodiment, the detectable label is a radiolabeled nucleotide.

[0131] In a further preferred embodiment, the experimental DNAs are labeled differentially from the control DNA, especially if both the DNA types are hybridized to the same microarray. The control DNA can comprise target polynucleotide molecules from normal individuals (i.e., those not afflicted with the neurological disorder or subjects who have not undergone to therapeutic treatment). In one preferred embodiment, the control DNA comprises target polynucleotide molecules pooled from samples from normal individuals. In one embodiment of the methods for generating a gene profile or DNA methylation profile of a therapeutic treatment, the control DNA is derived from the same subject, but taken at a different time point, such as before, during or after the therapeutic treatment.

[0132] Nucleic acid hybridization and wash conditions are chosen so that the DNA molecules

specifically bind or specifically hybridize to the complementary polynucleotide sequences of the array, preferably to a specific array site, wherein its complementary DNA is located. Arrays containing double-stranded probe DNA situated thereon are preferably subjected to denaturing conditions to render the DNA single-stranded prior to contacting with the DNA molecules. Arrays containing single-stranded probe DNA (e.g., synthetic oligodeoxyribonucleic acids) may need to be denatured prior to contacting with the DNA molecules. Optimal hybridization conditions will depend on the length (e.g., oligomer versus polynucleotide greater than 200 bases) and type (e.g., RNA, or DNA) of probe and target nucleic acids. One of skill in the art will appreciate that as the oligonucleotides become shorter, it may become necessary to adjust their length to achieve a relatively uniform melting temperature for satisfactory hybridization results. General parameters for specific (i.e., stringent) hybridization conditions for nucleic acids are described in Sambrook et al., *MOLECULAR CLONING--A LABORATORY MANUAL* (2ND ED.), Vols. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y. (1989), and in Ausubel et al., *CURRENT PROTOCOLS IN MOLECULAR BIOLOGY*, vol. 2, Current Protocols Publishing, New York (1994). Typical hybridization conditions for the cDNA microarrays of Schena et al. are hybridization in 5X SSC plus 0.2% SDS at 65° C for four hours, followed by washes at 25° C in low stringency wash buffer (1X SSC plus 0.2% SDS), followed by 10 minutes at 25° C in higher stringency wash buffer (0.1X SSC plus 0.2% SDS) (Schena et al., *Proc. Natl. Acad. Sci. U.S.A.* 93:10614 (1993)). Useful hybridization conditions are also provided in, e.g., Tijssen, 1993, *HYBRIDIZATION WITH NUCLEIC ACID PROBES*, Elsevier Science Publishers B. V.; and Kricka, 1992, *NONISOTOPIC DNA PROBE TECHNIQUES*, Academic Press, San Diego, Calif. Hybridization conditions may include hybridization at a temperature at or near the mean melting temperature of the probes (e.g., within 5° C., more preferably within 2° C) in 1 M NaCl, 50 mM MES buffer (pH 6.5), 0.5% sodium sarcosine and 30% formamide.

[0133] When fluorescently labeled DNAs are used in the aforementioned methods, the fluorescence emissions at each site of a microarray may be, preferably, detected by scanning confocal laser microscopy. In one embodiment, a separate scan, using the appropriate excitation line, is carried out for each of the two fluorophores used. Alternatively, a laser may be used that allows simultaneous specimen illumination at wavelengths specific to the two fluorophores and emissions from the two fluorophores can be analyzed simultaneously (see Shalon et al., 1996, "A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization," *Genome Research* 6:639-645, which is incorporated by reference in its entirety for all purposes). In one preferred embodiment, the arrays are scanned with a laser fluorescent scanner with a computer controlled X-Y stage and a microscope objective. Sequential excitation of the two fluorophores is achieved with a multi-line, mixed gas laser and the emitted light is split by wavelength and detected with two photomultiplier tubes. Fluorescence laser scanning devices are described in Schena et al., *Genome Res.* 6:639- 645 (1996), and in other references cited herein. Alternatively, the fiber-optic bundle described

by Ferguson et al., *Nature Biotech.* 14:1681-1684 (1996), may be used to monitor differentially DNA-methylated gene or DNA abundance levels at a large number of sites simultaneously.

[0134] Signals may be recorded and, in a preferred embodiment, analyzed by computer, e.g., using a 12 or 16 bit analog to digital board. In one embodiment the scanned image is despeckled using a graphics program (e.g., Hijaak Graphics Suite) and then analyzed using an image gridding program that creates a spreadsheet of the average hybridization at each wavelength at each site. If necessary, an experimentally determined correction for "cross talk" (or overlap) between the channels for the two fluors may be made. For any particular hybridization site on the transcript array, a ratio of the emission of the two fluorophores can be calculated. The ratio is independent of the absolute expression level of the differentially DNA-methylated gene, but is useful for differentially DNA-methylated genes whose expression is significantly modulated in association with the different neurological conditions.

[0135] In another embodiment of the present invention, changes in differentially DNA-methylated gene expression may be assayed in at least one cell of a subject by measuring transcriptional initiation, transcript stability, translation of transcript into protein product, protein stability, or a combination thereof. The gene, gene products, transcript, or polypeptide can be assayed by techniques such as in vitro transcription, quantitative nuclease protection assay (qNPA) analysis, focused gene chip analysis, Northern hybridization, nucleic acid hybridization, reverse transcription-polymerase chain reaction (RT-PCR), run-on transcription, Southern hybridization, electrophoretic mobility shift assay (EMSA), fluorescent or histochemical staining, microscopy and digital image analysis, and fluorescence activated cell analysis or sorting (FACS).

[0136] A reporter or selectable marker gene whose protein product is easily assayed may be used for convenient detection. Reporter genes include, for example, alkaline phosphatase, β -galactosidase (LacZ), chloramphenicol acetyltransferase (CAT), β -glucuronidase (GUS), bacterial/insect/marine invertebrate luciferases (LUC), green and red fluorescent proteins (GFP and RFP, respectively), horseradish peroxidase (HRP), β -lactamase, and derivatives thereof (e.g., blue EBFP, cyan ECFP, yellow-green EYFP, destabilized GFP variants, stabilized GFP variants, or fusion variants sold as LIVING COLORSTM fluorescent proteins by Clontech, Mountain View, CA). Reporter genes would use cognate substrates that are preferably assayed by a chromogen, fluorescent, or luminescent signal. Alternatively, assay product may be tagged with a heterologous epitope (e.g., FLAG, MYC, SV40 T antigen, glutathione transferase, hexahistidine, maltose binding protein) for which cognate antibodies or affinity resins are available.

[0137] In another embodiment, the gene or transcript can be assayed by using systems that employ expression vectors. An expression vector is a recombinant polynucleotide that is in chemical form either a deoxyribonucleic acid (DNA) and/or a ribonucleic acid (RNA). The physical form of the

expression vector may also vary in strandedness (e.g., single-stranded or double-stranded) and topology (e.g., linear or circular). The expression vector is preferably a double-stranded deoxyribonucleic acid (dsDNA) or is converted into a dsDNA after introduction into a cell (e.g., insertion of a retrovirus into a host genome as a provirus). The expression vector may include one or more regions from a mammalian gene expressed in the microvasculature, especially endothelial cells (e.g., ICAM-2, tie), or a virus (e.g., adenovirus, adeno-associated virus, cytomegalovirus, fowlpox virus, herpes simplex virus, lentivirus, Moloney leukemia virus, mouse mammary tumor virus, Rous sarcoma virus, SV40 virus, vaccinia virus), as well as regions suitable for genetic manipulation (e.g., selectable marker, linker with multiple recognition sites for restriction endonucleases, promoter for in vitro transcription, primer annealing sites for in vitro replication). The expression vector may be associated with proteins and other nucleic acids in a carrier (e.g., packaged in a viral particle) or condensed with chemicals (e.g., cationic polymers) to target entry into a cell or tissue.

[0138] The expression vector further comprises a regulatory region for gene expression (e.g., promoter, enhancer, silencer, splice donor and acceptor sites, polyadenylation signal, cellular localization sequence). Transcription can be regulated by tetracycline or dimerized macrolides. The expression vector may be further comprised of one or more splice donor and acceptor sites within an expressed region; Kozak consensus sequence upstream of an expressed region for initiation of translation; and downstream of an expressed region, multiple stop codons in the three forward reading frames to ensure termination of translation, one or more mRNA degradation signals, a termination of transcription signal, a polyadenylation signal, and a 3' cleavage signal. For expressed regions that do not contain an intron (e.g., a coding region from a cDNA), a pair of splice donor and acceptor sites may or may not be preferred. It would be useful, however, to include mRNA degradation signal(s) if it is desired to express one or more of the downstream regions only under the inducing condition. An origin of replication may also be included that allows replication of the expression vector integrated in the host genome or as an autonomously replicating episome. Centromere and telomere sequences can also be included for the purposes of chromosomal segregation and protecting chromosomal ends from shortening, respectively. Random or targeted integration into the host genome is more likely to ensure maintenance of the expression vector but episomes could be maintained by selective pressure or, alternatively, may be preferred for those applications in which the expression vector is present only transiently.

[0139] An expressed region may be derived from any gene of interest, and be provided in either orientation with respect to the promoter; the expressed region in the antisense orientation will be useful for making cRNA and antisense polynucleotide. The gene may be derived from the host cell or organism, from the same species thereof, or designed de novo; but it is preferably of archaeal, bacterial, fungal, plant, or animal origin. The gene may have a physiological function of one or more

nonexclusive classes: axon guidance, synaptic transmission or plasticity, myelination, long-term potentiation, neuron toxicity, embryonic development, regulation of actin networks, KEGG pathway, digestion, liver toxicity (hepatic stellate cell activation, fibrosis, and cholestasis), inflammation, oxidative stress, epilepsy, apoptosis, cell survival, differentiation, the unfolded protein response, Type II diabetes and insulin signaling, endocrine function, circadian rhythm, cholesterol metabolism and the steroidogenesis pathway, adhesion proteins; steroids, cytokines, hormones, and other regulators of cell growth, mitosis, meiosis, apoptosis, differentiation, circadian rhythm, or development; soluble or membrane receptors for such factors; adhesion molecules; cell-surface receptors and ligands thereof; cytoskeletal and extracellular matrix proteins; cluster differentiation (CD) antigens, antibody and T-cell antigen receptor chains, histocompatibility antigens, and other factors mediating specific recognition in immunity; chemokines, receptors thereof, and other factors involved in inflammation; enzymes producing lipid mediators of inflammation and regulators thereof; clotting and complement factors; ion channels and pumps; transporters and binding proteins; neurotransmitters, neurotrophic factors, and receptors thereof; cell cycle regulators, oncogenes, and tumor suppressors; other transducers or components of signaling pathways; proteases and inhibitors thereof; catabolic or metabolic enzymes, and regulators thereof. Some genes produce alternative transcripts, encode subunits that are assembled as homopolymers or heteropolymers, or produce propeptides that are activated by protease cleavage. The expressed region may encode a translational fusion; open reading frames of the regions encoding a polypeptide and at least one heterologous domain may be ligated in register. If a reporter or selectable marker is used as the heterologous domain, then expression of the fusion protein may be readily assayed or localized. The heterologous domain may be an affinity or epitope tag.

Methods of Identifying or Characterizing Therapeutic Compounds

[0140] Another aspect of the invention is identification or screening of chemical or genetic compounds, derivatives thereof, and compositions including same that are effective in treatment of neurological diseases or disorders and individuals at risk thereof. The amount that is administered to an individual in need of therapy or prophylaxis, its formulation, and the timing and route of delivery is effective to reduce the number or severity of symptoms, to slow or limit progression of symptoms, to inhibit expression of one or more of the aforementioned differentially DNA-methylated genes that are transcribed at a higher level in neurological disease, to activate expression of one or more of the aforementioned differentially DNA-methylated genes that are transcribed at a lower level in neurological disease, or any combination thereof. Determination of such amounts, formulations, and timing and route of drug delivery is within the skill of persons conducting in vitro assays, in vivo studies of animal models, and human clinical trials.

[0141] A screening method may comprise administering a candidate compound to an organism or

incubating a candidate compound with a cell, and then determining whether or not gene expression is modulated. Such modulation may be an increase or decrease in activity that partially or fully compensates for a change that is associated with or may cause neurological disease. Differentially DNA-methylated gene expression may be increased at the level of rate of transcriptional initiation, rate of transcriptional elongation, stability of transcript, translation of transcript, rate of translational initiation, rate of translational elongation, stability of protein, rate of protein folding, proportion of protein in active conformation, functional efficiency of protein (e.g., activation or repression of transcription), or combinations thereof. See, for example, U.S. Patent Numbers 5,071,773 and 5,262,300. High-throughput screening assays are possible (e.g., by using parallel processing and/or robotics).

[0142] The screening method may comprise incubating a candidate compound with a cell containing a reporter construct, the reporter construct comprising transcription regulatory region covalently linked in a cis configuration to a downstream gene encoding an assayable product; and measuring production of the assayable product. A candidate compound which increases production of the assayable product would be identified as an agent which activates gene or cDNA expression while a candidate compound which decreases production of the assayable product would be identified as an agent which inhibits gene or cDNA expression. See, for example, U.S. Patent Numbers 5,849,493 and 5,863,733.

[0143] The screening method may comprise measuring in vitro transcription from a reporter construct in the presence or absence of a candidate compound (the reporter construct comprising a transcription regulatory region) and then determining whether transcription is altered by the presence of the candidate compound. In vitro transcription may be assayed using a cell-free extract, partially purified fractions of the cell, purified transcription factors or RNA polymerase, or combinations thereof. See, for example, U.S. Patent Numbers 5,453,362, 5,534,410, 5,563,036, 5,637,686, 5,708,158 and 5,710,025.

[0144] Techniques for measuring transcriptional or translational activity in vivo are known in the art. For example, a nuclear run-on assay may be employed to measure transcription of a reporter gene. Translation of the reporter gene may be measured by determining the activity of the translation product. The activity of a reporter gene can be measured by determining one or more of transcription of polynucleotide product (e.g., RT-PCR of GFP transcripts), translation of polypeptide product (e.g., immunoassay of GFP protein), and enzymatic activity of the reporter protein per se (e.g., fluorescence of GFP or energy transfer thereof).

[0145] Another aspect of the invention provides methods of identifying, or predicting the efficacy of, test compounds. In particular, the invention provides methods of identifying compounds which mimic the effects of behavioral therapies. In still another aspect, the systems and methods described herein

provide a method for predicting efficacy of a test compound for altering a behavioral response, by obtaining a database, e.g., as described in greater detail above, treating a test animal or human (e.g., a control animal or human that has not undergone other therapies, such as behavioral therapy) with the test compound, and comparing genomic or cDNA expression data of tissue samples from the animal or human treated with the test compound to measure a degree of similarity with one or more differentially DNA-methylated gene profiles in said database. In certain embodiments, the untreated animal or human exhibits a psychological and/or behavioral abnormality possessed by the animals or humans used to generate the database prior to administration of the behavioral therapy.

[0146] In another aspect of the invention, a method is provided for predicting efficacy of a test compound for altering a behavioral response in a subject with at least one autism spectrum disorder comprising: (a) preparing a microarray comprising a plurality of different oligonucleotides, wherein the oligonucleotides are specific to differentially DNA-methylated genes associated with an autism spectrum disorder; (b) obtaining a differentially DNA-methylated gene profile representative of the differentially DNA-methylated gene profile of at least one sample of a selected tissue type from a subject subjected to each of at least one of a plurality of selected behavioral therapies which promote the behavioral response; (c) administering the test compound to the subject; and (d) comparing the differentially DNA-methylated gene profile data in at least one sample of the selected tissue type from the subject treated with the test compound to determine a degree of similarity with one or more differentially DNA-methylated gene profiles associated with an autism spectrum disorder; wherein the predicted efficacy of the test compound for altering the behavioral response is correlated to said degree of similarity.

[0147] In another aspect, the systems and methods described herein relate to methods of identifying small molecules useful for treating neurological conditions.

[0148] For example, in another embodiment a database of differentially DNA-methylated gene profile data representative of the genetic expression response of a selected neuronal tissue type from an animal that was subjected to at least one of a plurality of behavioral therapies and that has undergone a selected physiological change since commencement of the behavioral therapy may be obtained. In an exemplary embodiment, subjects (e.g., subjects that display a preselected behavioral abnormality, such as an autism spectrum disorder neurological condition (including for example autistic disorder, pervasive developmental disorder-not otherwise specified (PDD-NOS), including atypical autism, Asperger's Disorder, Rett's syndrome), Parkinson's disease, parkinsonism, cognitive impairments, age-associated memory impairments, cognitive impairments, dementia associated with neurologic and/or neurological conditions, allodynia, catalepsy, hypernociception, and epilepsy, brain tumors, brain lesions, multiple sclerosis, Down's syndrome, progressive supranuclear palsy, frontal lobe syndrome,

schizophrenia, delirium, Tourette's syndrome, myasthenia gravis, attention deficit hyperactivity disorder, dyslexia, mania, depression, apathy, myopathy, Alzheimer's disease, Huntington's Disease, dementia, encephalopathy, schizophrenia, severe clinical depression, brain injury, Attention Deficit Disorder (ADD), Attention Deficit Hyperactivity Disorder (ADHD), hyperactivity disorder, bipolar manic-depressive disorder, ischemia, alcohol addiction, drug addiction, obsessive compulsive disorders, Pick's disease and Binswanger's disease or a combination thereof), are subjected to behavioral therapy (including, for example, applied behavior analysis (ABA) intervention methods, dietary changes, exercise, massage therapy, group therapy, talk therapy, play therapy, conditioning, or alternative therapies such as sensory integration and auditory integration therapies), and their tissues (including, for example, and not by way of limitation, lymphocytes, blood, or mucosal epithelial cells, brain, spinal cord, heart, arteries, esophagus, stomach, small intestine, large intestine, liver, pancreas, lungs, kidney, urinary tract, ovaries, breasts, uterus, testis, penis, colon, prostate, bone, muscle, cartilage, thyroid gland, adrenal gland, pituitary, bone marrow, blood, thymus, spleen, lymph nodes, skin, eye, ear, nose, teeth or tongue, and/or neurological tissues (including, for example, and not by way of limitation, olfactory bulb cells, cerebrospinal fluid, hypothalamus, amygdala, pituitary, nervous system, brainstem, cerebellum, cortex, frontal cortex, hippocampus, striatum, and thalamus) or a combination thereof are examined for physiological changes (one or more improvements in social interaction, language abilities, restricted interests, repetitive behaviors, sleep disorders, seizures, gastrointestinal, hepatic, and mitochondrial function, neural inflammation, or a combination thereof), and differentially DNA methylation status and/or genetic expression responses are obtained for tissues that have undergone a desired change. In certain embodiments, the subjects are further selected for having undergone a desired change in behavior as well.

[0149] From such a database, biological targets for intervention can be identified, such as potential therapeutics (e.g., genes or cDNAs that are upregulated and thus may exert a beneficial effect on the physiology and/or behavior of the subject), potential receptor targets (e.g., receptors associated with upregulated proteins, the activation of which receptors may exert a beneficial effect on the physiology and/or behavior of the subject; or receptors associated with downregulated proteins, the inhibition of which may exert a beneficial effect on the physiology and/or behavior of the subject). In certain embodiments, one or more genes or one or more cDNAs, the expression of which differs by a statistically significant amount in a treated subject as compared to an untreated control, may be selected as targets for intervention.

[0150] Small molecule test agents may then be screened in any of a number of assays to identify those with potential therapeutic applications. The term "small molecule" refers to a compound having a molecular weight less than about 2500 amu, preferably less than about 2000 amu, even more preferably less than about 1500 amu, still more preferably less than about 1000 amu, or most preferably less than

about 750 amu. For example, subjects or tissue samples may be treated with such test agents to identify those that produce similar changes in expression of the targets, or produce similar gene profiles or DNA methylation profiles, as can be obtained by administration of behavioral therapy. Alternatively or additionally, such test agents may be screened against one or more target receptors to identify compounds that agonize or antagonize these receptors, singly or in combination, e.g., so as to reproduce or mimic the effect of behavioral therapy.

[0151] Compounds that induce a desired effect on targets, tissue, or subjects may then be selected for clinical development, and may be subjected to further testing, e.g., therapeutic profiling, such as testing for efficacy and toxicity in subjects. Analogs of selected compounds, e.g., compounds having similar cores but varying substituents and stereochemistry, may similarly be developed and tested. Agents that have acceptable characteristics for therapeutic use in humans or animals may be prepared as pharmaceutical preparations, e.g., with a pharmaceutically acceptable excipient (such as a non-pyrogenic or sterile excipient). Such agents may also be licensed to a manufacturer for development and/or commercialization, e.g., for manufacture and sale of a pharmaceutical preparation comprising said selected agent.

[0152] Accordingly, one aspect of the invention provides a method for predicting efficacy of a test compound for altering a behavioral response in a subject with at least one autism spectrum disorder comprising: (a) preparing a microarray comprising a plurality of different oligonucleotides, wherein the oligonucleotides are specific to differentially DNA-methylated genes associated with an autism spectrum disorder; (b) obtaining a differentially DNA-methylated gene profile representative of the differentially DNA-methylated gene profile of at least one sample of a selected tissue type from a subject subjected to each of at least one of a plurality of selected behavioral therapies which promote the behavioral response; (c) administering the test compound to the subject; and (d) comparing differentially DNA-methylated gene profile data in at least one sample of the selected tissue type from the subject treated with the test compound to determine a degree of similarity with one or more differentially DNA-methylated gene profiles associated with an autism spectrum disorder; wherein the predicted efficacy of the test compound for altering the behavioral response is correlated to said degree of similarity.

[0153] In one embodiment of the foregoing methods, step (a) comprises obtaining a differentially DNA-methylated gene profile representative of the differentially DNA-methylated gene profile of at least two samples of a selected tissue type referred to supra. In a related embodiment, step (a) comprises obtaining a differentially DNA-methylated gene profile data representative of the differentially DNA-methylated gene profile of at least three samples of a selected tissue referred to supra. In one embodiment in which the more than one sample of a selected tissue type referred to supra

is used to determine a differentially DNA-methylated gene profile, the selected tissue types are different tissue types, whereas in other embodiments the tissue types are the same. For example, in an exemplary embodiment, a tissue type may be lymphoblastoid cells and a second tissue type olfactory bulb cells, such that the differentially DNA-methylated gene profile generated from these two tissue samples in the treated subject may be compared to the differentially DNA-methylated gene profiles derived from the subjects subjected to the behavioral therapy. In other embodiments, differentially DNA-methylated gene profiles may be generated from multiple samples of the same tissue type from the same animal, such as blood samples taken at different intervals during the behavioral therapy.

[0154] In another embodiment of the foregoing methods, the differentially DNA-methylated gene profile is that shown in Table 1, Table 2, Table 3, or a combination thereof. In another embodiment, the differentially DNA-methylated gene comprises at least 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95% or 98% of the differentially DNA-methylated genes shown in Table 1, Table 2, Table 3, or a combination thereof. In another embodiment, the differentially DNA-methylated gene profile comprises at least 5, 10, 15, 20, 25 or 30 of differentially DNA-methylated genes listed in Table 1, Table 2, Table 3, or a combination thereof. In yet another embodiment of the foregoing methods, the differentially DNA-methylated gene profile comprises an increase or a decrease in expression of at least one to 73 or any integer value thereof, at least one to 201 or any integer value thereof, or at least one to 50 or any integer value thereof, of any of the genes listed in Table 1, Table 2, or Table 3, respectively, or a combination thereof.

[0155] In one embodiment of the foregoing methods, the selected tissue type comprises a neuronal tissue type, such as a neuronal tissue type selected from the group consisting of olfactory bulb cells, cerebrospinal fluid, hypothalamus, amygdala, pituitary, nervous system, brainstem, cerebellum, cortex, frontal cortex, hippocampus, striatum, and thalamus. In another embodiment, the selected tissue type is selected from the group consisting of brain, spinal cord, heart, arteries, esophagus, stomach, small intestine, large intestine, liver, pancreas, lungs, kidney, urinary tract, ovaries, breasts, uterus, testis, penis, colon, prostate, bone, muscle, cartilage, thyroid gland, adrenal gland, pituitary, bone marrow, blood, thymus, spleen, lymph nodes, skin, eye, ear, nose, teeth and tongue.

[0156] In one embodiment, the behavioral therapy comprises applied behavior analysis (ABA) intervention methods, dietary changes, exercise, massage therapy, group therapy, talk therapy, play therapy, conditioning, or alternative therapies such as sensory integration and auditory integration therapies.

[0157] In one embodiment of the foregoing methods, the test subject or animal is a human. In another embodiment, the animal is a non-human animal. Such non-human animals include vertebrates such as rodents, non-human primates, ovines, bovines, ruminants, lagomorphs, porcines, caprines, equines,

canines, felines, aves, etc. Preferred non-human animals are selected from the order Rodentia, most preferably mice. The term "order Rodentia" refers to rodents (i.e., placental mammals (Class Euthria) which include the family Muridae (rats and mice). In a specific embodiment, the test animal is a mammal, a primate, a rodent, a mouse, a rat, a guinea pig, a rabbit or a human.

[0158] The test compound may be administered to the subject or animal using any mode of administration, including, intravenous, subcutaneous, intramuscular, intrastemal, topical, liposome-mediate, rectal, intravaginal, ophthalmic, intracranial, intraspinal or intraorbital. The test compound may be administered once or more than once as part of a treatment regimen. In some embodiments, additional test compounds or agents may be administered to the subject animal to ascertain the efficacy of the test compound or the combination of test compounds or agents. In some embodiments, a differentially DNA-methylated gene profile may also be obtained from the subject or animal prior to treatment with the test agent. In such embodiments, the efficacy of the test agent may be determined by comparing the differentially DNA-methylated gene profile of the subject or animal after treatment with the compound with (a) the differentially DNA-methylated gene profile prior to treatment with the compound and (b) to the differentially DNA-methylated gene profile for the behavioral therapy. For example, if the test compound causes the differentially DNA-methylated gene profile to approach that of said differentially DNA-methylated gene profile, the test compound may be predicted to be efficacious.

[0159] It is understood by one skilled in the art that the order of steps (a) and (b) in the foregoing methods may be interchanged i.e. the subject or animal may be treated with the compound prior to obtaining the differentially DNA-methylated gene profile for the behavior therapy. Accordingly, the invention also provides a method wherein step (b) is performed prior to step (a).

[0160] When comparing the gene differentially DNA-methylated gene profile data in at least one sample of the selected tissue type from the subject or animal treated with the test compound to determine a degree of similarity with one or more differentially DNA-methylated gene profiles, any number of statistical methods known to one skilled in the art may be used. In some embodiments, a differentially DNA-methylated gene profile may be obtained from samples of a test subject or animal prior to the administration of the test compound or from a control subject or animal to generate a control differentially DNA-methylated gene profile for each of the tissue types of interest. In such embodiments, the differentially DNA-methylated gene profile from the tissue types of the test subjects or animal(s) may be compared to both the control differentially DNA-methylated gene profiles and the differentially DNA-methylated gene profiles resulting from the behavioral therapy to determine to which of these differentially DNA-methylated gene profiles the differentially DNA-methylated gene profile is most similar. If they are more similar to the control differentially DNA-methylated gene

profile, the test compound may be considered less efficacious, whereas if it is more similar to the differentially DNA-methylated gene profile of the behavioral therapy then the compound is considered more efficacious.

[0161] In one variation of the methods described herein, more than one test compound may be administered to the test subject or animal, such that the efficacy of a combination of test compounds is tested. In another variation, rather than using, or in addition to using, a test compound, a nonchemical test agent is also applied to the subject or animal, such as for example, and not by way of limitation, temperature, humidity, sunlight exposure or any other environmental factor. In yet another environment, the subject or animal is subjected to an invasive or noninvasive surgical procedure, in lieu or in addition to the test compound. In such embodiments, the efficacy of the surgical procedure may be ascertained.

[0162] In still yet another aspect, the systems and methods described herein relate to a kit for identifying a compound for treating a behavioral disorder, comprising a database, e.g., as described in greater detail above, and a computer program for comparing differentially DNA-methylated gene profile data obtained from assays wherein a test compound is administered to an untreated subject or animal with differentially DNA-methylated gene profile data in the database and identifying similarity between the differentially DNA-methylated gene profile data from the assays and one or more stored differentially DNA-methylated gene profiles.

[0163] In yet another aspect of the invention, the systems and methods described herein relate a kit is provided for identifying a compound for treating at least one autism spectrum disorder comprising (a) a database having information stored therein one or more differentially DNA-methylated gene profiles specific for the differentially DNA-methylated gene listed in Table 1, Table 2, Table 3, or a combination thereof, of subjects that have been subjected to at least one of a plurality of selected autism spectrum disorder neurological therapies and wherein the subject has undergone a desired physiological change; and (b) a computer program for comparing differentially DNA-methylated gene profile data obtained from assays wherein a test compound is administered to a subject with the database and providing information representative of a measure of similarity between the differentially DNA-methylated gene profile data and one or more stored differentially DNA-methylated gene profiles.

[0164] Another aspect of the invention provides a method of assessing treatment efficacy in an individual having a neurological disorder comprising determining the expression level of one or more of the aforementioned informative differentially DNA-methylated genes in Table 1, Table 2, Table 3, or a combination thereof at multiple time points during treatment, wherein (i) a decrease in expression of the one or more informative differentially DNA-methylated genes shown to be expressed at

increased levels in individuals having a neurological disorder or at risk for developing a neurological disorder as compared with a control, or (ii) an increase in expression of the one or more informative differentially DNA-methylated genes shown to be expressed at decreased levels in individuals having a neurological disorder or at risk for developing a neurological disorder as compared with a control, or both (i) and (ii), is indicative that treatment is effective.

[0165] The invention also provides a method of assessing the efficacy of a treatment in an individual having a neurological disorder, comprising (i) determining differentially DNA-methylated gene profile data in a plurality of patient samples, obtained at multiple time points during treatment of the patient, of a selected tissue type; (ii) determining a degree of similarity between (a) the differentially DNA-methylated gene profile data in the patient samples; and (b) a differentially DNA-methylated gene profile produced by a therapy which has been shown to be efficacious in treatment of the neurological disorder; wherein a high degree of similarity is indicative that the treatment is effective. Preferably, the high degree of similarity is a log₂ ratio of less than +/- 0.3.

[0166] In one embodiment, the invention also provides a method for assessing the efficacy of a treatment in an individual having at least one autism spectrum disorder comprising (a) determining differentially DNA-methylated gene profile data specific for at least one of the differentially DNA-methylated genes set out in Table 1 or Table 2 or a combination thereof, in a plurality of patient samples of a selected tissue type from a patient receiving the treatment; (b) determining a degree of similarity between (i) the differentially DNA-methylated gene profile data in the patient samples; and (ii) a differentially DNA-methylated gene specific for the differentially DNA-methylated genes set out in listed in Table 1, Table 2, Table 3, or a combination thereof, produced by a therapy which has been shown to be efficacious in treatment of the at least one autism spectrum disorder; wherein a high degree of similarity between the differentially DNA-methylated gene profile data of (ii) with the differentially DNA-methylated gene profile data of (a) during or after treatment is indicative that the treatment is effective. Preferably the high degree of similarity is a log₂ ratio of less than +/- 0.3.

[0167] Another aspect of the invention provides kits. One aspect provides a kit for identifying a compound for treating a behavioral or neurological disorder, comprising (i) a database having information stored therein differentially DNA-methylated gene profile data representative of the differentially DNA-methylated gene expression response of selected tissue type samples from subjects or animals that have been subjected to at least one of a plurality of selected behavioral therapies and wherein the tissue has undergone a desired physiological change; and (ii) a computer program for (a) comparing differentially DNA-methylated gene profile data obtained from assays, where a test compound is administered to a subject or an animal, with the database; and (b) providing information representative of a measure of similarity between the differentially DNA-methylated gene profile data

and one or more stored profiles.

[0168] In yet another aspect of the invention, a kit is provided for identifying a compound for treating at least one autism spectrum disorder comprising (a) a database having information stored therein one or more differentially DNA-methylated gene profiles specific for the differentially DNA-methylated genes listed in Table 1, Table 2, Table 3, or a combination thereof, of subjects that have been subjected to at least one of a plurality of selected autism spectrum disorder neurological therapies and wherein the subject has undergone a desired physiological change; and (b) a computer program for comparing differentially DNA-methylated gene profile data obtained from assays wherein a test compound is administered to a subject with the database and providing information representative of a measure of similarity between the differentially DNA-methylated gene profile data and one or more stored gene profiles or DNA methylation profiles.

[0169] In some embodiments of the methods described herein, the test compound comprises an antibody or fragment thereof, a nucleic acid molecule, antisense reagent, a small molecule drug, or a nutritional or herbal supplement. Test compounds can be screened individually, in combination with one or more other compounds, or as a library of compounds. In one embodiment, test compounds include nucleic acids, peptides, polypeptides, peptidomimetics, RNAi constructs, antisense oligonucleotides, ribozymes, antibodies, small molecules, and nutritional or herbal supplements or a combination thereof.

[0170] In general, test compounds for modulation of neurological disorders, including those autistic spectrum disorders such as autistic disorder, pervasive developmental disorder-not otherwise specified (PDD-NOS), including atypical autism, Asperger's Disorder, or a combination thereof, can be identified from large libraries of natural products or synthetic (or semi-synthetic) extracts or chemical libraries according to methods known in the art. Those skilled in the field of drug discovery and development will understand that the precise source of test extracts or compounds is not critical to the screening procedure(s) of the invention. Accordingly, virtually any number of chemical extracts or compounds can be screened using the exemplary methods described herein. Examples of such extracts or compounds include, but are not limited to, plant-, fungal-, prokaryotic- or animal-based extracts, fermentation broths, and synthetic compounds, as well as modification of existing compounds. Numerous methods are also available for generating random or directed synthesis (e.g., semi-synthesis or total synthesis) of any number of chemical compounds, including, but not limited to, saccharide-, lipid-, peptide-, and nucleic acid-based compounds. Synthetic compound libraries are commercially available, e.g., Chembridge (ChemBridge Corporation, San Diego, Calif.). Alternatively, libraries of natural compounds in the form of bacterial, fungal, plant, and animal extracts are commercially available from a number of sources, including Biotics (Sussex, UK), Xenova (Slough, UK), Harbor

Branch Oceanographics Institute (Ft. Pierce, Fla.), and PharmaMar, U.S.A. (Cambridge, Mass.). In addition, natural and synthetically produced libraries are generated, if desired, according to methods known in the art, e.g., by standard extraction and fractionation methods. Furthermore, if desired, any library or compound is readily modified using standard chemical, physical, or biochemical methods.

Methods of Conducting Drug Discovery

[0171] Another aspect of the invention provides methods for conducting drug discovery related to the methods and differentially DNA-methylated gene chips provided herein.

[0172] One aspect of the invention provides a method for conducting drug discovery comprising: (a) generating a database of differentially DNA-methylated gene profile data representative of the genetic expression response of at least one selected tissue type (for example, one of the aforementioned neuronal tissue types) from a subject or an animal that was subjected to at least one of a plurality of behavioral therapies and that has undergone a selected physiological change since commencement of the behavioral therapy; (b) selecting at least one differentially DNA-methylated gene profile from Table 1, Table 2, Table 3, or a combination thereof and selecting at least one target as a function of the selected differentially DNA-methylated gene profiles; (c) screening a plurality of small molecule test agents in assays to obtain differentially DNA-methylated gene profile data associated with administration of the agents and comparing the obtained data with the one or more selected differentially DNA-methylated gene profiles; (d) selecting for clinical development test agents that exhibit a desired effect on the target as evidenced by the differentially DNA-methylated gene profiles data; (e) for test agents selected for clinical development, conducting therapeutic profiling of the test compound, or analogs thereof, for efficacy and toxicity in subjects or animals; and (f) selecting at least one test agent that has an acceptable therapeutic and/or toxicity profile.

[0173] Another aspect of the invention provides a method for conducting drug discovery comprising: (a) generating a database of differentially DNA-methylated gene profile data representative of the genetic expression response of at least one selected neuronal tissue type from a subject or an animal that was subjected to at least one of a plurality of behavioral therapies and that has undergone a selected physiological change since commencement of the behavioral therapy; (b) administering small molecule test agents to test subjects or animals to obtain differentially DNA-methylated gene profile data associated with administration of the agents and comparing the obtained data with the one or more selected differentially DNA-methylated gene profiles; (c) selecting test agents that induce profiles similar to profiles obtainable by administration of behavioral therapy; (d) conducting therapeutic profiling of the selected test compound(s), or analogs thereof, for efficacy and toxicity in subjects or animals; and (e) identifying a pharmaceutical preparation including one or more agents identified in step (e) as having an acceptable therapeutic and/or toxicity profile.

[0174] In one embodiment, the database of differentially DNA-methylated gene profile data representative of the genetic expression response of at least one selected neuronal tissue type from a subject or an animal that was subjected to at least one of a plurality of behavioral therapies and that has undergone a selected physiological change since commencement of the behavioral therapy comprises at least one differentially DNA-methylated gene profile from Table 1, Table 2, Table 3, or a combination thereof

EXAMPLES

[0175] The invention now being generally described, it will be more readily understood by reference to the following examples, which are included merely for purposes of illustration of certain aspects and embodiments of the present invention, and are not intended to limit the invention, as one skilled in the art would recognize from the teachings hereinabove and the following examples, that other DNA microarrays, neurological conditions, cognitive therapies or data analysis methods, all without limitation, can be employed, without departing from the scope of the invention as claimed. The contents of any patents, patent applications, patent publications, or scientific articles referenced anywhere in this application are herein incorporated in their entirety.

EXAMPLE 1

[0176] Global methylation profiling of lymphoblastoid cell lines reveals epigenetic contributions to autism spectrum disorders and a novel autism candidate gene, RORA, whose protein product is reduced in autistic brain

[0177] In this example global methylation profiling of discordantly diagnosed monozygotic twins and their nonautistic siblings on CpG island arrays is used to test the hypothesis that differential gene expression in idiopathic autism is, at least in part, the result of aberrant methylation. The study reveals distinct methylation differences in multiple genes between the discordant MZ twins as well as common epigenetic differences distinguishing the twins (the undiagnosed twin exhibiting milder autistic traits that are below the threshold for diagnosis) from nonautistic sibling controls. Overlaying methylation profiles with gene expression profiles previously obtained for these individuals (18) revealed specific differentially expressed ASD candidate genes which are potential targets of aberrant methylation. Immunohistochemical staining of the protein products of two of the candidate genes (Bcl-2 and RORA) which showed increased methylation in LCL revealed protein reduction in autistic cerebellum and frontal cortex as postulated, demonstrating the usefulness of LCL as a surrogate model with which to explore mechanisms of dysregulation of gene expression in ASD.

Materials and Methods

Cell culture and 5-Aza-2-deoxycytidine treatment

[0178] LCL, which were obtained from the Autism Genetic Research Exchange (AGRE; Los Angeles, CA) were derived from lymphocytes of autistic and normal siblings and cultured as previously described (Hu et al. (2006) BMC Genomics 7, 118). Demethylating treatments were carried out by incubating cells in culture with 5 μ M 5-Aza-2-deoxycytidine (Fluka, St. Louis, MO) or vehicle control, DMSO (FisherBioTech, Pittsburgh, PA) for 48 hours before harvesting cells.

DNA and RNA isolation procedures

[0179] Genomic DNA was isolated using the Qiagen DNeasyTM Blood and Tissue Kit (Valencia, CA) and used for bisulfite modification for methylation analyses. Total RNA was isolated using TRIzol reagent (Invitrogen, Carlsbad, CA) according to the manufacturer's protocol and used for gene expression analyses by quantitative RT-PCR. The RNA samples were further purified using the RNeasy Mini kit (Qiagen, Valencia, CA) and tested for integrity on RNA 6000 NanoChips using an Agilent 2100 Bioanalyzer.

Selection of samples

[0180] Three pairs of male monozygotic twins discordant for diagnosis of autism as determined by the Autism Diagnostic Interview-Revised (ADI-R) assessment instrument were selected for this study. The undiagnosed co-twin from each set, though not clinically autistic, exhibited some autistic traits and were described as either "Broad spectrum" or "Not quite autistic (NQA)" by the AGRE repository based on their stated criteria (www.agre.org). Cell lines derived from respective non-autistic siblings of two pairs of twins were also included in the analyses, in addition to cell lines derived from a set of monozygotic twins unaffected by autism. For confirmation experiments (qRT-PCR, bisulfite sequencing, and methylation-specific PCR), two additional case-control pairs of siblings (affected autistic individual and respective unaffected male siblings) were analyzed. LCL from individuals with specific genetic and chromosomal abnormalities, diagnosed co-morbid disorders, or those born prematurely (<35 weeks of gestation) were excluded from this study.

Global methylation profiling using CpG island microarrays

[0181] Global methylation profiling on CpG island arrays was performed as previously described by Rauch et al. (Rauch et al. (2006) Cancer Res 66, 7939-47) to screen for differentially methylated autism candidate genes. Briefly, 20 μ g of genomic DNA was digested with the frequent cutter, MseI (5'- TTAA). Two unidirectional linkers: oligo-Long (5'-GCGGTGACCCGGGAGATCTGAATTC-

3')(SEQ ID NO: 1) and oligo-MseI (5'- TAGAATTCAGATCTCCCG-3') (SEQ ID NO:2) were annealed and ligated to the MseI digested genomic DNA by incubation at 16°C overnight. Enrichment of the methylated fraction was accomplished by methyl-CpG-island recovery assay (MIRA). Briefly, 1 µg of purified GST-tagged MBD2b protein and 1 µg of purified His-tagged MBD3L1 protein were pre-incubated with glutathione beads in the presence of SC110 genomic DNA at 4°C on a rocking platform. The beads were then incubated with 1 µg of the digested and linker-ligated genomic DNA in a binding reaction mixture [10mmol/L Tris-HCl (pH 7.5), 50 mmol/L NaCl, 1 mmol/L EDTA, 1 mmol/L DTT, 3 mmol/L MgCl₂, 0.1% Triton-X100, 5% glycerol, 25 µg/mL bovine serum albumin, and 1.25 µg/mL (dcm minus) SC110 bacterial DNA]. The methylation-enriched MIRA fraction was washed three times with wash buffer [10mM Tris-HCl (pH 7.5), 700mM NaCl, 1mM EDTA, 3mM MgCl₂, 0.1% Triton-X-100], eluted from the glutathione beads using the Qiagen QIAquick™ PCR purification kit (Valencia, CA) and PCR-amplified. Enriched and unenriched fractions derived from each individual were then indirectly labeled with aminoallyl-dUTP, conjugated with either Alexa 647 or Alexa 555 (Invitrogen, Carlsbad, CA) and hybridized onto 8.1K CpG island microarrays (UHN Microarray Centre, University of Toronto, Canada). For all paired analyses, a direct comparison was performed in which the methylation-enriched fractions from two individuals were pooled and hybridized onto the same microarray. In addition, indirect comparisons were performed by co-hybridizing the methylation-enriched (MIRA) fraction with the respective unenriched DNA fraction obtained from the same individual. For each paired analysis (between autistic MZ twins and/or between autistic co-twin and unaffected sibling), a total number of 4 replicates were performed, including direct and indirect comparisons.

Microarray data analysis

[0182] Following hybridization and washing according to established protocols, the CpG island microarray slides were scanned using an Axon GenePix™ 4000B laser scanner to elicit dye intensities for each element on the array. Analysis of raw intensity data was performed using the TIGR TM4 Software Suite (Quackenbush, J. (2002) Microarray data normalization and transformation. *Nature Genetics* 32, 496-501). Raw intensity data were normalized and filtered using LOWESS normalization and standard deviation (variance) regularization programs contained within the MIDAS module of TM4. The normalized intensities were then uploaded into the MeV module and a 70% data filter was applied to remove CpG island regions for which log₂ intensity values are missing in >30% of the samples before statistical analyses of the data. Differentially methylated CpG islands were identified using the Significance Analysis of Microarrays (SAM) (Tusher et al. (2001) *Proc Natl Acad Sci USA* 98, 5116-21) module within MeV, which also provided false discovery rates (FDR) for each analysis. A one-class SAM analysis identified differential methylation between discordant monozygotic twins as well as across both twins in comparison to unaffected control siblings.

[0183] To identify genes that were differentially expressed in these samples, gene expression data from our earlier expression profiling of LCL from these subjects (Hu et al. (2006) *BMC Genomics* 7, 118) were downloaded from the Gene Expression Omnibus database (GEO Accession # GSE4187) and analyzed using a 70% data filter and a one-class SAM across all sib pairs. This analysis resulted in the identification of a total of 3043 differentially expressed genes with a FDR of 10%. Microsoft Access software was used to identify the overlap between differentially methylated and differentially expressed genes between the twins and normal siblings, using Gene Symbol as the common identifier between datasets. The log2 values indicating relative expression or methylation for the twins vs. the normal siblings were obtained by averaging the values across all replicates.

Quantitative PCR analyses (qRT-PCR)

[0184] Two candidate genes, BCL-2 and RORA, were chosen for quantitative real-time (RT)-PCR analysis to examine the correlation between expression of mRNA transcript and methylation status as predicted by microarray analysis, as well as to examine the effect of 5-Aza-2-deoxycytidine treatment on transcript levels. Total RNA was reverse transcribed into cDNA using the iScript cDNA Synthesis Kit (Bio-Rad, Hercules, CA) following the manufacturer's protocol, and qPCR was accomplished using Invitrogen's Platinum SYBR Green qPCR SuperMix-UDG with ROX. The qPCR reactions (in triplicate) were analyzed on an ABI Prism 7300 Sequence Detection System and transcript levels of BCL-2 and RORA were quantified and normalized to standard curves obtained using universal 18S rRNA primers. Sequences of primers used for qRT-PCR are given in Table 4.

Bisulfite modification, bisulfite sequencing and methylation-specific PCR (MSP)

[0185] Two ug of genomic DNA isolated from LCL were treated with sodium bisulfite using the Active Motif MethylDetector kit (Carlsbad, CA) according to the manufacturer's specifications. Successful bisulfite treatment and conversion was verified using positive control PCR primers for the p16 (CDKN2A) locus provided in the MethylDetector kit.

[0186] For direct bisulfite sequencing of the BCL-2 promoter region, bisulfite-modified DNA was used for a nested PCR amplifying the P1 promoter region of BCL-2, which overlapped with the CpG island region (UHNhscpg0002696; chr18:59137439-59137855) contained on the 8.1K CpG island microarray. Successful PCR amplification was verified by gel electrophoresis. PCR products were extracted from the gel and purified using the Qiagen Qiaquick Gel Extraction kit (Valencia, CA) according to the manufacturer's specifications. Purified PCR products were directly sequenced using the Applied Biosystems Automated 3730xl DNA Analyzer (Cornell DNA Sequencing Facility, Ithaca, NY) with the reverse primer to obtain methylation status of CpG dinucleotides.

[0187] For methylation-specific PCR (MSP), 2 sets of primers spanning a region approximately 300

bp upstream of the first exon of RORA (UHNhscpg0014103; chr15:59309160-59309428) were designed according to assumed differences in methylation status (i.e., fully methylated or fully unmethylated). These primers were previously successful in identifying methylation-specific silencing of RORA in several gastric cancer cell lines (Yamashita et al. (2006) *Cancer Science* 97, 64-71). Bisulfite-modified genomic DNA was used as a template for PCR and PCR products were resolved by agarose gel electrophoresis. Sequences of primers used for bisulfite sequencing analysis of BCL-2 and MSP of RORA are given in Table 4.

Immunohistochemistry

[0188] Two different sets of tissue arrays were obtained through the Autism Tissue Program (San Diego, CA). One set of arrays was prepared by Dr. Charles Eberhart in the Johns Hopkins Pathology Department Tissue Microarray Core Facility, as described previously (Eberhart et al. (2006) *Journal of Autism and Developmental Disorders* 36, 1131-1135). The samples included thin sections from the cerebellar cortex of 5 autistic, 1 Rett individual and 5 age-matched control individuals. Additional clinical data on each case has been previously reported by Eberhart et al. (Eberhart et al. (2006)). The second set of postmortem tissue arrays are from Dr. Janine LaSalle's laboratory and contained slices from the BA9 region of the frontal cortex from control and age- and sex-matched autistic individuals as well as individuals with a variety of developmental disorders as described by Nagarajan et al. (Nagarajan et al. (2006) *Epigenetics* 1, 172-182).

[0189] Immunohistochemistry for RORA and BCL-2 protein was performed using a standard avidin-biotin complex method of detection described by Eberhart et al. (Eberhart et al. (2006)). Briefly, tissue array slides were deparaffinized in xylene and rehydrated using a dilution series of 100%, 95% and 75% ethanol. Slides were subjected to heat-induced antigen retrieval by incubation in Tris-EDTA buffer (10mM Tris Base, 1mM EDTA, 0.05% Tween 20, pH 9.0) at 95°C for 25 minutes. Slides were then incubated overnight at 4°C with a rabbit polyclonal primary antibody (1:50 dilution) against RORA (4ug/mL, Santa Cruz Biotechnology: sc-28612) or against BCL-2 (4ug/mL, Santa Cruz Biotechnology: sc-492). Biotinylated secondary goat anti-rabbit antibody (1:200 dilution) was added to slides for 30 minutes at room temperature. All incubations were conducted in a humidified chamber. Antibodies were detected using the Santa Cruz rabbit ABC Staining System using an avidin-biotin complex method and diaminobenzadine (DAB) as a chromagen. Fixed slides were imaged using an Olympus BX-60 upright light microscope.

Pathway and functional analyses of differentially methylated/expressed genes

[0190] Differentially methylated (or expressed) genes were uploaded into either Ingenuity Pathway Analysis (IPA) (Ingenuity Systems, Inc. Redwood City, CA) or Pathway Studio 5 (PS5)(Ariadne Genomics Inc., Rockville MD) to discover relationships between the genes as well as their association

with various processes and disorders. The Fisher Exact Test was used to identify significant functions and pathways represented within the respective datasets, with the p value indicating the probability that the stated function or process is not associated with the genes in the dataset.

Results

[0191] Global methylation profiling of LCL from discordant monozygotic twins and unaffected siblings reveals differential methylation of genes relevant to neurological functions and disorders.

[0192] CpG island arrays were used to identify differentially methylated genes from LCL of monozygotic twins discordant in diagnosis of autism. Methylation-enriched DNA was isolated from genomic DNA using a pull-down method called methylated CpG island recovery assay (MIRA), previously established by Rauch et al. (Rauch et al. (2006) Cancer Res 66, 7939-47). Methylation-enriched and unenriched DNA fractions were then hybridized onto CpG island microarrays to identify CpG islands with differential methylation patterns between individuals. The CpG island microarray analysis identified 73 CpG islands as differentially methylated between discordant monozygotic twins [Table 1]. Network analyses of the differentially methylated genes immediately at the 3'-end, or overlapping, the CpG islands revealed enrichment of high level functions biologically relevant to the autism phenotype, and included numerous genes involved in neurological disorders as well as nervous system development and function (Table 5). In fact, although only 1 twin of each pair is diagnosed autistic, his co-twin still has autistic traits (which we call "mild"), but falls below the criteria for being diagnosed autistic. Thus, if the combined CpG island data from both co-twins are compared against their respective normal sibs, there are a larger number (201) of differentially methylated CpG islands because the phenotypes involved in the comparisons are distinctly different, even though the genotypes are similar (but not identical) since the twins and their respective nonautistic sibling are genetically related [Table 2]. Network analysis of this set of differentially methylated genes revealed enrichment for high level functions shown in Table 6. Interestingly, the function with the highest statistical significance ($p = 9.02 \times 10^{-8}$ by the Fisher Exact test) includes genes involved in transcriptional control of gene expression, many of which are required for proper brain development. Other significant high level functions represented within the gene dataset are nervous system development and function, neurological disease, and cell death.

Overlap of differentially methylated genes with differentially expressed genes

[0193] To prioritize ASD candidate genes for further analysis, differentially methylated genes from the analysis of twins vs. normal siblings were compared to differentially expressed genes derived from prior gene expression profiling of these same samples on DNA microarrays (Hu et al. 2006). Table 3 lists the genes that are both differentially expressed and methylated, and Figure 1A displays the inverse relationship between expression level and methylation level for all but one of these genes. Network

analysis was then performed to examine the relationship between this set of genes and biological processes. As shown in Figure 1B, many of the associated processes within the network, including synaptic regulation, fetal development, morphogenesis, apoptosis, inflammation, digestion, steroid biosynthesis, and mental deficiency, have been associated with autism. Two genes from this network, BCL-2 and RORA, were selected for further study because of their respective roles in apoptosis and morphogenesis/inflammation. Interestingly, BCL-2 protein has been demonstrated to be reduced in the cerebellum and frontal cortex of autistic subjects relative to control subjects, but RORA, a nuclear steroid hormone receptor and transcriptional activator that is involved in Purkinje cell differentiation (Boukhtouche et al. (2006) *Cerebellum* 5, 97-104) and cerebellar development (Gold et al. (2007) *Brain Research* 1140, 19-25), has never before been implicated in autism. In addition, RORA, a regulator of circadian rhythm, is also neuroprotective against inflammation and oxidative stress.

Confirmation of differential methylation of specific CpG sites in upstream promoter regions of BCL-2 and RORA.

[0194] The CpG islands immediately upstream of BCL-2 and RORA were chosen for further confirmation of methylation status in LCL of autistic and control individuals. Specifically, a 417bp CpG island overlapping the P1 promoter region of the BCL-2 gene (chr18:59137439-59137855) was identified as hypermethylated in all co-twins exhibiting autistic traits compared to unaffected siblings in the CpG island microarray, and a 269bp CpG island associated with the 5' end of the RORA1 isoform (chr15:59309160- 59309428) was identified as hypermethylated in the same analysis. To obtain high-resolution information regarding which specific CpG sites may be hyper-methylated in autistic individuals, 2 separate experimental approaches taking advantage of bisulfite modification of genomic DNA were employed: bisulfite sequencing and methylation-specific PCR. Bisulfite modification of genomic DNA results in deamination of all unmethylated cytosine residues to uracils, while methylated cytosine residues are protected against conversion.

[0195] Figure 2 displays the results from bisulfite sequencing of the BCL-2 promoter region after bisulfite treatment of genomic DNA isolated from LCL derived from the monozygotic twins, their unaffected siblings, an additional autistic/unaffected sib pair (A_2020/C_2019), and a pair of unaffected monozygotic twins (C_2744/C_2745). Following bisulfite conversion, nested primers were designed to amplify across the P1 promoter region of BCL-2 (Fig. 2A). This is the same region interrogated by the CpG island microarrays and spans 38 total CpG sites. As shown in Figure 2B, an overall increase in % methylation across the 38 CpG sites (~25% on average) was observed in the diagnosed autistic samples in comparison to an average of 11.4% for the undiagnosed co-twins and 9.5% for the unaffected siblings. Increased methylation was especially noted on CpG sites #3-13 for autistic individuals only. Also, there appears to be a decrease in methylation at site 34 in the autistic samples that is found methylated in 4 out of 5 control individuals. Interestingly, the same methylated

CpG sites are not observed across all autistic subjects (with the exception of CpG site #12), suggesting that regulation is not dependent upon specific sites, but instead may involve the entire region encompassed by CpG sites #3-13.

[0196] Methylation-specific PCR (MSP) was utilized to analyze methylation in a region upstream of RORA because of the presence of homopolymer sequences between the CpG sites of interest which would complicate sequencing analyses. This region overlaps with the region interrogated by the CpG island microarray, and is located approximately 300 bp before the first exon of the ROR α 1 isoform (Fig. 3A). Two sets of primers encompassing several CpG sites were designed according to assumed differences in methylation status to amplify the same region upstream of RORA. These particular primers were previously successful in identifying epigenetic silencing of RORA in several gastric cancer cell lines (Yamashita et al. (2006) *Cancer Science* 97, 64-71). The presence of a PCR product corresponding to methylated DNA only in the vehicle-treated samples of autistic individuals in contrast to those of the unaffected controls (Fig. 3B) confirms methylation of the RORA promoter region only in the autistic individuals. The absence of methylated product in the autistic samples following treatment with 5 μ M 5-Aza-2-deoxycytidine further confirms methylation in the untreated autistic samples and specificity of the methylated PCR primers.

Confirmation of methylation-dependent silencing of candidate genes identified by CpG island arrays

[0197] Quantitative RT-PCR was used to confirm decreased expression of BCL-2 and RORA in autistic samples and to evaluate the effect of a global methylation inhibitor, 5-Aza-2-deoxycytidine, on gene expression. For both BCL-2 and the RORA, gene expression was significantly higher ($p < 0.05$) in the unaffected control than autistic co-twins (Fig 4A). Generally, the diagnosed autistic co-twin (A) had the lowest level of expression of BCL-2 and RORA, while the milder undiagnosed co-twin (M) exhibited transcript levels between that observed for unaffected sibling controls and autistic co-twins. This suggests a quantitative relationship between phenotype and gene expression of these two genes, although additional studies are required to confirm this observation. BCL-2 and RORA gene expression was found to be methylation-dependent as both transcripts were up-regulated by 1.5- and 1.8-fold, respectively, following global inhibition of methylation with 5-Aza-2-deoxycytidine (Fig. 4B). By contrast, there were no statistically significant increases in the expression levels of BCL-2 and RORA in the LCL from the "mild" co-twins with 5-Aza-2-deoxycytidine treatment.

RORA and BCL-2 proteins are also reduced in post-mortem brain tissues of autistic subjects

[0198] To explore the relevance of our findings of enhanced methylation and reduced expression of RORA and BCL-2 in LCL to the expression of these proteins in the brain, the levels of RORA and BCL-2 protein were investigated in post-mortem tissues from autistic subjects and age- and sex-

matched controls, using tissue arrays containing sections from both the cerebellum and frontal cortex (Yamashita et al. (2006) *Cancer Science* 97, 64-71; Fatemi et al. (2001) *J Autism Dev Disord* 31, 529-35). Immunohistochemical staining of the tissue arrays confirmed reduced levels of RORA protein in the cerebellum of autistic individuals, including the molecular cell, granule cell and Purkinje cell layers, where expression of RORA has been previously demonstrated (Gold et al. (2007) *Brain Research* 1140, 19-25; Hadj-Sahraoui et al. (2001) *Developmental Brain Research* 126, 201-209; Doulazmi et al. (1999) *Journal of Comparative Neurology* 411, 267-273). Representative immunohistochemical staining results were obtained for RORA protein in cerebellum of autistic subjects in comparison to that of respective age- and sex-matched controls (data not shown). Immunoreactivity was decreased in most autistic samples (male: A, C, I and female: E, G) as indicated by the decreased extent of RORA staining. As reduced expression of RORA (as well as other regulators of the circadian rhythm) was observed only in the subtype of ASD exhibiting severe language impairment (Hu et al. (2009) *Autism Research* 2, 78-97), reduced RORA protein in all autistic brain samples are not expected.

[0199] Immunohistochemistry of post-mortem brain tissue also revealed decreased BCL-2 protein in the granule cell and Purkinje cell layers of the cerebellum, with 3 out of 5 autistic samples showing reduced immunoreactivity for BCL-2 protein (data not shown). Higher magnification representative images of the cerebellar sections show that both RORA and BCL-2 protein are specifically reduced in Purkinje cells (Fig. 5). RORA protein was observed in both the cytoplasm and nucleus, as reported previously for different isoforms of RORA (Giguere et al. (1994) *Genes and Development* 8, 538-553), while counterstaining with hematoxylin revealed that BCL-2 is confined to the cytoplasm. Figure 6 shows that RORA is also reduced in the majority autistic tissue sections vs. age- and sex-matched control sections from the frontal cortex (BA9 region), suggesting that RORA deficiency is not limited to the cerebellum.

[0200] In this study, we used microarray-based epigenomic profiling of LCL from discordant monozygotic autistic twins and unaffected siblings to test the hypothesis that DNA methylation is involved in global dysregulation of gene expression in ASD and to screen for candidate genes whose expression is potentially linked to aberrant DNA methylation. We observed DNA-methylation differences in numerous loci containing genes important to the pathobiology of autism. In particular, functional and pathway analyses of the differentially methylated/expressed genes showed enrichment of genes involved in inflammation and apoptosis, cellular differentiation, brain morphogenesis, growth rate, cytokine production, myelination, synaptic regulation, learning, and steroid biosynthesis, all of which were altered in ASD. The candidate genes were prioritized for further analyses by identifying the overlap between the differentially methylated genes and those that had been shown to be differentially expressed in the same set of samples in previous gene expression analyses. Pathway

analyses of this filtered set of genes thus focused our attention on 2 genes, BCL-2 and RORA, as candidate genes for ASD whose expression were dysregulated by aberrant methylation.

[0201] In this study, global methylation profiling of LCL from phenotypically discordant monozygotic twins and nonautistic siblings highlighted the role of epigenetic regulation in idiopathic autism and reveals DNA methylation as a mechanism through which this regulation may occur. Two candidate genes, BCL-2 and RORA, which were identified from our coupling of methylation with gene expression data, exhibited increased methylation of specific CpG sites in respective upstream regulatory CpG islands that coincided with methylation-specific gene silencing. Translation of these findings from LCL to the detection of decreased BCL-2 and RORA protein in post-mortem brain tissues of autistic individuals further confirms the feasibility of using LCL as a surrogate model for autism, particularly when investigating dysregulated genes with systemic functions, such as apoptosis and circadian rhythm. In addition to identifying key autism candidate genes, these studies also yield further insight into the pathobiology of this complex disorder by elucidating global epigenomic modifications relevant to the autistic phenotype. Aside from shedding light on higher order regulation of gene expression, another compelling reason to investigate epigenetic mechanisms in idiopathic autism is that such modifications can be influenced by exposure to biological modulators and environmental factors. Epigenetics may thus mediate the interaction between genotype and intrinsic (biologic) and extrinsic (environmental) factors contributing to ASD.

BCL-2 as a candidate gene for autism

[0202] BCL-2 is an anti-apoptotic protein located in the outer mitochondrial membrane that is important for cell survival under a variety of stressful conditions. Our data provides evidence of the involvement of reduced BCL-2 in ASD. Moreover, our study provides new evidence that increased DNA methylation of the P1 promoter region of BCL-2 is a possible mechanism for increased gene silencing and decreased protein abundance of BCL-2 in the autistic brain. These studies also localized the changes in BCL-2 to Purkinje cells in the cerebellum.

RORA is a novel candidate gene for autism

[0203] RORA is a member of the NR1 subfamily of nuclear hormone receptors and is involved in transcriptional regulation of many genes. Among its many relevant functions are regulation of the circadian clock by activation of BMAL1, neuroprotection in the face of oxidative stress and inflammation, survival and differentiation of Purkinje cells, and cerebellar development. Multiple regulators of circadian rhythm, including RORA, were aberrantly expressed in LCL from autistic individuals with severe language impairment. We have confirmed the methylation and expression differences of RORA in LCL, and the involvement of RORA in autistic brain pathology.

[0204] As shown in Figs. 3 and 4 respectively, RORA was confirmed to be inversely differentially methylated and expressed in LCL from autistic vs. nonautistic siblings, with expression dependent on methylation, as demonstrated by the absence of methylation in the presence of 5-Aza-2-deoxycytidine. We have also shown by immunohistochemical staining of cerebellar and frontal cortex regions of autistic vs. normal brain (Fig. 6), that RORA protein is noticeably reduced in the majority of the autistic samples relative to age- and sex-matched controls. This reduction was also specifically demonstrated in Purkinje cells which are dependent on RORA for both survival and differentiation (Fig. 5). These findings thus linked molecular changes identified in a peripheral cell model of ASD to actual pathological changes in the autistic brain, suggesting that LCL was an appropriate surrogate for studies on autism.

Role of DNA methylation as an epigenetic mechanism for dysregulation of gene expression in autism

[0205] The broad functions of the methylation-regulated differentially expressed genes identified in this study provided support for our findings that the pathological symptoms of autism were in part the result of systemic dysregulation of gene expression. Bioinformatic analysis of the differentially methylated genes from the CpG island assays (Tables 5 and 6) revealed enrichment for many high level functions implicated in autism including aberrant response to inflammation, decreased cell survival, and abnormal cellular differentiation leading to abnormal brain development. In addition, these processes reiterated common biological themes in autism that our laboratory has identified from global gene expression profiling of LCL in discordant monozygotic twins (Hu et al. 2006), sibling pairs (Hu et al. (2009) *PLoS ONE* 4, e5775), and large scale case-control studies of unrelated subjects (Hu et al. (2009) *Autism Research* 2, 78-97).

[0206] Coupling methylation profiles with global gene expression profiles of the same autistic and control individuals provided a means of identifying high-priority candidate genes for autism. Confirmation of the postulated decrease in protein products of two of the candidate genes in post-mortem brain tissue of autistic individuals highlights the relevance of our findings in LCL to the autistic brain. This ability to translate findings from peripheral tissues to the brain is consistent with reports of individuals with neuropsychiatric disorders having detectable epigenetic programming disturbances, both in the brain and in secondary tissues. LCL and primary blood lymphocytes have been demonstrated to be reasonable surrogate experimental models of autism to identify both expressed and epigenetic alterations that may also have diagnostic potential as biomarkers for autism (see, e.g., Baron et al. (2006) *J Autism Dev Disord* 36, 973-82; Nishimura et al. (2007) *Hum Mol Genet* 16, 1682-98; Gregg et al. (2008) *Genomics* 91, 22-29, and; Enstrom et al. (2009) *Brain, Behavior, and Immunity* 23, 124-133).

Table 1

Differentially methylated CpG islands between discordant monozygotic twins. 73 CpG islands were found to be significantly differentially methylated between discordant monozygotic twins using SAM analysis (FDR = 23%). Each CpG island is associated with up to three genes. These genes are directly overlapping, immediately upstream of, or immediately downstream of the CpG island interrogated. Relative methylation of the CpG island is expressed as the average log 2 ratio of the hybridization signals for each MZ twin comparison. UHNid = unique identifier used by University Health Network Microarray Centre.

UHNID	Genome Location of CpG island	Gene overlapping CpG site	Upstream Gene	Distance upstream	Downstream Gene	Distance downstream	Relative methylation log ₂ (A/M)
UHNhscpg0000030	chr2:44076267-44077277	LRPPRC	ABCG8	117159	PPM1B	172226	-0.367
UHNhscpg0000124	chr2:70967886-70969007		CD207	51425	VAX2	12220	-0.456
UHNhscpg0000131	chr6:24827829-24828199		C6orf62	447	GMNN	54943	-0.327
UHNhscpg0000159	chr1:199408487-199408821		TMEM9	16235	DKFZp434B1231	41571	-0.283
UHNhscpg0000250	chr6:106535603-106535953		PREP	577941	PRDM1	104934	-0.314
UHNhscpg0000316	chr4:26817110-26817285		STIM2	181009	FLJ45721	939	-0.438
UHNhscpg0000545	chr1:7936516-7937311		TNFRSF9	13042	PARK7	7068	-0.266
UHNhscpg0000546	chr21:42056430-42056937	RIPK4	C21orf129	47426	PRDM15	34518	-0.222
UHNhscpg0000615	chr22:26502436-26502794	MN1	AK124820	1103819	PITPNB	74863	-0.299
UHNhscpg0000759	chr9:99498610-99499528	XPA	NCBP1	24878	FOXE1	155829	-0.311
UHNhscpg00003095	chr10:84007638-84007793		SH2D4B	1611342	GHITM	1881498	-0.290
UHNhscpg00004204	chr17:35826867-35828053	TOP2A	RARA	59447	IGFBP4	25148	-0.385
UHNhscpg00004536	chr4:154113407-154113462	KIAA1727	ARFIP1	60896	TRIM2	231605	-0.379
UHNhscpg00004802	chr9:131138996-131139140	AK092192	C9orf106	11991	AK130729	152600	-0.309
UHNhscpg00004936	chr21:46567171-46567396	C21orf58	C21orf58	4672	PCNT	1086	-0.326
UHNhscpg00005059	chr21:31851621-31852410	TIAM1	AB096964	518955	SOD1	101395	-0.426
UHNhscpg00005064	chr2:26249810-26250309		RAB10	36028	HADHA	16703	-0.350
UHNhscpg00005174	chr8:86302636-86302733	E2F5	LRRCC1	57564	BC032347	13910	-0.482
UHNhscpg00005396	chr17:38178680-38179548	VPS25	RAMP2	10100	VNKA	6673	-0.215
UHNhscpg00005784	chr14:90930199-90930271		AB040942	80391	KIAA2010	63437	-0.429
UHNhscpg00005914	chr19:54648171-54648348	ALDH16A1	FLJ20643	1292	FLT3LG	20949	-0.300
UHNhscpg00005966	chr12:16561287-16561466		MGST1	152677	LMO3	31107	-0.229
UHNhscpg00006260	chr18:34166142-34166247		BRUNOL4	766201	PIK3C3	3622949	-0.439
UHNhscpg00007310	chr2:169477583-169478200		G6PC2	2832	ABCB11	9494	-0.398
UHNhscpg00008604	chr21:28375719-28375880		C21orf94	58576	C21orf100	457630	-0.428

UHNhscpg0009207	chr19:51611501-51611950	CCDC8	2820	FLJ10781	49638	-0.505
UHNhscpg0009771	chr5:84194219-84194473	EDIL3	477852	BC050328	1419544	-0.385
UHNhscpg0011090	chr13:98870327-98870739	PHGDHL1	33638	TM9SF2	80989	-0.375
UHNhscpg0013020	chr5:37244700-37245302	FLJ13231	25195	NUP155	82395	-0.270
UHNhscpg0013713	chr7:135312 312915	UNQ1940	228111	CHRM2	891023	-0.458
UHNhscpg0013730	chr20:25156 56512	ENTPD6	678	PYGB	20193	-0.331
UHNhscpg0013793	chr12:53097940-53098512	ZNF385	33192	FAM112B	37499	-0.437
UHNhscpg0013846	chr1:10413353-10413818	PGD	10566	CORT	18739	-0.416
UHNhscpg0013977	chr11:458511-459622	AK128239	6439	AK125720	5568	-0.238
UHNhscpg0014040	chr19:12640945-12641575	MAN2B1	2117	MORG1	8	-0.536
UHNhscpg0014051	chr20:49007736-49008634	ADNP	26802	MOC53	135	-0.291
UHNhscpg0014093	chr14:103441860-103442558	PPP1R13B	58305	C14orf2	5821	-0.473
UHNhscpg0014095	chr19:6313403-6313686	ASAH3	28841	ALKBH7	9757	-0.350
UHNhscpg0014096	chr17:76887872-76888250	MGC15523	4181	C17orf55	2968	-0.197
UHNhscpg0014153	chr11:126090589-126091362	ST3GAL4	274627	ETS1	1742509	-0.384
UHNhscpg0014295	chr19:42032423-42032894	AK000357	11324	ZNF345	212	-0.316
UHNhscpg0014384	chr8:98287780-98288369	PGCP	62883	TSPYL5	66521	-0.311
UHNhscpg0014470	chr19:6201401-6202238	ACSBG2	57289	ASAH3	55486	-0.284
UHNhscpg0014588	chr15:35175550-35176026	MEIS2	199017	TMCO5	838092	-0.328
UHNhscpg0014721	chr7:37348386-37349330	ELMO1	137672	GPR141	397190	-0.504
UHNhscpg0014724	chr11:27340902-27341886	CCDC34	234972	LGR4	2199	-0.313
UHNhscpg0014982	chr9:116389648-116390322	ATP6V1G1	82097	C9orf91	23138	-0.291
UHNhscpg0015416	chr17:18905950-18906317	GRAP	14889	AK026312	30694	-0.349
UHNhscpg0015475	chr20:51631246-51632258	AK024093	5992	BC065723	292299	-0.459
UHNhscpg0015493	chr12:108822867-108823464	GLTP	20191	GIT2	28527	-0.199
UHNhscpg0015585	chr13:65054119-65054708	AK098560	1737860	PCDH9	720259	-0.346
UHNhscpg0015971	chr17:75370070-75370624	ENPP7	39455	CBX8	12146	-0.193
UHNhscpg0016169	chr4:46667541-46667774	COX7B2	78661	GABRB1	60561	-0.183
UHNhscpg0016274	chr1:39110922-39111736	RRAGC	12995	GJA10	589	-0.423
UHNhscpg0016365	chr8:87030008-87030494	REXO1L1	120023	ATP6V0D2	149760	-0.613
UHNhscpg0016460	chr14:69825628-69826013	SLC8A3	100088	C14orf112	36524	-0.200
UHNhscpg0017302	chr14:54807341-54808469	DLG7	79192	KIAA0831	94393	-0.491
UHNhscpg0017317	chr10:119297732-119298170	PDZD8	172805	RAB11FIP2	456248	-0.332
UHNhscpg0017365	chr17:35047459-35047834	PPP1R1B	1058	TCAP	27290	-0.265
UHNhscpg0017365	chr17:35047459-35047834	STARD3				
UHNhscpg0017365	chr17:35047459-35047834	FBXO34				
UHNhscpg0017365	chr17:35047459-35047834	EMX2				
UHNhscpg0017365	chr17:35047459-35047834	ATP6V1G1				
UHNhscpg0017365	chr17:35047459-35047834	CBX2				
UHNhscpg0017365	chr17:35047459-35047834	GABRA4				
UHNhscpg0017365	chr17:35047459-35047834	MYCBP				
UHNhscpg0017365	chr17:35047459-35047834	REXO1L1				
UHNhscpg0017365	chr17:35047459-35047834	SLC8A3				
UHNhscpg0017365	chr17:35047459-35047834	DLG7				
UHNhscpg0017365	chr17:35047459-35047834	PDZD8				
UHNhscpg0017365	chr17:35047459-35047834	PPP1R1B				
UHNhscpg0017365	chr17:35047459-35047834	STARD3				
UHNhscpg0017365	chr17:35047459-35047834	FBXO34				
UHNhscpg0017365	chr17:35047459-35047834	EMX2				
UHNhscpg0017365	chr17:35047459-35047834	ATP6V1G1				
UHNhscpg0017365	chr17:35047459-35047834	CBX2				
UHNhscpg0017365	chr17:35047459-35047834	GABRA4				
UHNhscpg0017365	chr17:35047459-35047834	MYCBP				
UHNhscpg0017365	chr17:35047459-35047834	REXO1L1				
UHNhscpg0017365	chr17:35047459-35047834	SLC8A3				
UHNhscpg0017365	chr17:35047459-35047834	DLG7				
UHNhscpg0017365	chr17:35047459-35047834	PDZD8				
UHNhscpg0017365	chr17:35047459-35047834	PPP1R1B				
UHNhscpg0017365	chr17:35047459-35047834	STARD3				
UHNhscpg0017365	chr17:35047459-35047834	FBXO34				
UHNhscpg0017365	chr17:35047459-35047834	EMX2				
UHNhscpg0017365	chr17:35047459-35047834	ATP6V1G1				
UHNhscpg0017365	chr17:35047459-35047834	CBX2				
UHNhscpg0017365	chr17:35047459-35047834	GABRA4				
UHNhscpg0017365	chr17:35047459-35047834	MYCBP				
UHNhscpg0017365	chr17:35047459-35047834	REXO1L1				
UHNhscpg0017365	chr17:35047459-35047834	SLC8A3				
UHNhscpg0017365	chr17:35047459-35047834	DLG7				
UHNhscpg0017365	chr17:35047459-35047834	PDZD8				
UHNhscpg0017365	chr17:35047459-35047834	PPP1R1B				
UHNhscpg0017365	chr17:35047459-35047834	STARD3				
UHNhscpg0017365	chr17:35047459-35047834	FBXO34				
UHNhscpg0017365	chr17:35047459-35047834	EMX2				
UHNhscpg0017365	chr17:35047459-35047834	ATP6V1G1				
UHNhscpg0017365	chr17:35047459-35047834	CBX2				
UHNhscpg0017365	chr17:35047459-35047834	GABRA4				
UHNhscpg0017365	chr17:35047459-35047834	MYCBP				
UHNhscpg0017365	chr17:35047459-35047834	REXO1L1				
UHNhscpg0017365	chr17:35047459-35047834	SLC8A3				
UHNhscpg0017365	chr17:35047459-35047834	DLG7				
UHNhscpg0017365	chr17:35047459-35047834	PDZD8				
UHNhscpg0017365	chr17:35047459-35047834	PPP1R1B				
UHNhscpg0017365	chr17:35047459-35047834	STARD3				
UHNhscpg0017365	chr17:35047459-35047834	FBXO34				
UHNhscpg0017365	chr17:35047459-35047834	EMX2				
UHNhscpg0017365	chr17:35047459-35047834	ATP6V1G1				
UHNhscpg0017365	chr17:35047459-35047834	CBX2				
UHNhscpg0017365	chr17:35047459-35047834	GABRA4				
UHNhscpg0017365	chr17:35047459-35047834	MYCBP				
UHNhscpg0017365	chr17:35047459-35047834	REXO1L1				
UHNhscpg0017365	chr17:35047459-35047834	SLC8A3				
UHNhscpg0017365	chr17:35047459-35047834	DLG7				
UHNhscpg0017365	chr17:35047459-35047834	PDZD8				
UHNhscpg0017365	chr17:35047459-35047834	PPP1R1B				
UHNhscpg0017365	chr17:35047459-35047834	STARD3				
UHNhscpg0017365	chr17:35047459-35047834	FBXO34				
UHNhscpg0017365	chr17:35047459-35047834	EMX2				
UHNhscpg0017365	chr17:35047459-35047834	ATP6V1G1				
UHNhscpg0017365	chr17:35047459-35047834	CBX2				
UHNhscpg0017365	chr17:35047459-35047834	GABRA4				
UHNhscpg0017365	chr17:35047459-35047834	MYCBP				
UHNhscpg0017365	chr17:35047459-35047834	REXO1L1				
UHNhscpg0017365	chr17:35047459-35047834	SLC8A3				
UHNhscpg0017365	chr17:35047459-35047834	DLG7				
UHNhscpg0017365	chr17:35047459-35047834	PDZD8				
UHNhscpg0017365	chr17:35047459-35047834	PPP1R1B				
UHNhscpg0017365	chr17:35047459-35047834	STARD3				
UHNhscpg0017365	chr17:35047459-35047834	FBXO34				
UHNhscpg0017365	chr17:35047459-35047834	EMX2				
UHNhscpg0017365	chr17:35047459-35047834	ATP6V1G1				
UHNhscpg0017365	chr17:35047459-35047834	CBX2				
UHNhscpg0017365	chr17:35047459-35047834	GABRA4				
UHNhscpg0017365	chr17:35047459-35047834	MYCBP				
UHNhscpg0017365	chr17:35047459-35047834	REXO1L1				
UHNhscpg0017365	chr17:35047459-35047834	SLC8A3				
UHNhscpg0017365	chr17:35047459-35047834	DLG7				
UHNhscpg0017365	chr17:35047459-35047834	PDZD8				
UHNhscpg0017365	chr17:35047459-35047834	PPP1R1B				
UHNhscpg0017365	chr17:35047459-35047834	STARD3				
UHNhscpg0017365	chr17:35047459-35047834	FBXO34				
UHNhscpg0017365	chr17:35047459-35047834	EMX2				
UHNhscpg0017365	chr17:35047459-35047834	ATP6V1G1				
UHNhscpg0017365	chr17:35047459-35047834	CBX2				
UHNhscpg0017365	chr17:35047459-35047834	GABRA4				
UHNhscpg0017365	chr17:35047459-35047834	MYCBP				
UHNhscpg0017365	chr17:35047459-35047834	REXO1L1				
UHNhscpg0017365	chr17:35047459-35047834	SLC8A3				
UHNhscpg0017365	chr17:35047459-35047834	DLG7				
UHNhscpg0017365	chr17:35047459-35047834	PDZD8				
UHNhscpg0017365	chr17:35047459-35047834	PPP1R1B				
UHNhscpg0017365	chr17:35047459-35047834	STARD3				
UHNhscpg0017365	chr17:35047459-35047834	FBXO34				
UHNhscpg0017365	chr17:35047459-35047834	EMX2				
UHNhscpg0017365	chr17:35047459-35047834	ATP6V1G1				
UHNhscpg0017365	chr17:35047459-35047834	CBX2				
UHNhscpg0017365	chr17:35047459-35047834	GABRA4				
UHNhscpg0017365	chr17:35047459-35047834	MYCBP				
UHNhscpg0017365	chr17:35047459-35047834	REXO1L1				
UHNhscpg0017365	chr17:35047459-35047834	SLC8A3				
UHNhscpg0017365	chr17:35047459-35047834	DLG7				
UHNhscpg0017365	chr17:35047459-35047834	PDZD8				
UHNhscpg0017365	chr17:35047459-35047834	PPP1R1B				
UHNhscpg0017365	chr17:35047459-35047834	STARD3				
UHNhscpg0017365	chr17:35047459-35047834	FBXO34				
UHNhscpg0017365	chr17:35047459-35047834	EMX2				
UHNhscpg0017365	chr17:35047459-35047834	ATP6V1G1				
UHNhscpg0017365	chr17:35047459-35047834	CBX2				
UHNhscpg0017365	chr17:35047459-35047834	GABRA4				
UHNhscpg0017365	chr17:35047459-35047834	MYCBP				
UHNhscpg0017365	chr17:35047459-35047834	REXO1L1				
UHNhscpg0017365	chr17:35047459-35047834	SLC8A3				
UHNhscpg0017365	chr17:35047459-35047834	DLG7				
UHNhscpg0017365	chr17:35047459-35047834	PDZD8				
UHNhscpg0017365	chr17:35047459-35047834	PPP1R1B				
UHNhscpg0017365	chr17:35047459-35047834	STARD3				

UHNhscpg0017459	chr10:96056036-96056426	PLCE1	BC068549	344749	NOC3L	26552	-0.188
UHNhscpg0017564	chr12:14847117-14848131	WBP11	H2AFJ	27894	ART4	25380	-0.407
UHNhscpg0017660	chr21:44887276-44887749	C21orf29	KRTAP10-10	4478	KRTAP10-11	3009	-0.301
UHNhscpg0017725	chr3:180805205-180805651	NDUFB5	MRPL47	77	USP13	47983	-0.407
UHNhscpg0017768	chr3:101721999-101722586	TMEM45A	TOMM70A	119084	GPR128	88548	-0.552
UHNhscpg0017855	chr4:26695450-26695704		STIM2	59349	FLJ45721	122520	-0.575
UHNhscpg0018235	chr14:88753793-88754227	CHES1	CHES1	59882	AK130927	28430	-0.587
UHNhscpg0019001	chr5:4348162-4348505		IRX1	693646	ADAMTS16	844937	-0.240
UHNhscpg0019014	chr4:158361-1361804	GRIA2	GLRB	49260	C4orf18	903378	-0.217
UHNhscpg0019155	chr4:7038747-7038826	TBC1D14	AF338194	7454	AK126065	43855	-0.371
UHNhscpg0019168	chr5:72565874-72566264		MGC13034	59150	BTF3	263741	-0.318
UHNhscpg0020409	chr8:91729041-91729327		TMEM64	1929	EFCBP1	143856	-0.318
UHNhscpg0020414	chr18:57372244-57373069	CDH20	MC4R	1181622	RNF152	258513	-0.292
UHNhscpg0020863	chr6:52968276-52968740		GSTA4	141	ICK	5316	-0.201

Table 2

Differentially methylated CpG islands between all twins and normal siblings. 201 CpG islands were found to be significantly differentially methylated between monozygotic twins and unaffected siblings as identified by SAM analysis (FDR = 10%). Each CpG island is associated with up to three genes. These genes are directly overlapping, immediately upstream of, or immediately downstream of the CpG island interrogated. Relative methylation of the CpG island is expressed as the average log₂ ratio of the hybridization signals for each co-twin and his respective unaffected sibling. UHNid = unique identifier used by University Health Network Microarray Centre.

UHNid	Genome Location of CpG island	Gene overlapping CpG island	Upstream Gene	Downstream	Relative methylation Δ Log ₂ (T/N)
UHNhscpg0000113	chr11:85243289-85244177	AK129942	SYTL2	MGC34732	-0.499
UHNhscpg0000124	chr2:70967886-70969007		CD207	VAX2	0.279
UHNhscpg0000144	chr8:75395030-75395910	JPH1	LY96	GDAP1	0.187
UHNhscpg0000250	chr6:106535603-106535953		PREP	PRDM1	0.117
UHNhscpg0000272	chr5:131853022-131854258	IRF1	LOC441108	IL5	0.054
UHNhscpg0000378	chr11:6581864-6582603	ILK	KIAA0409	TAF10	0.111
UHNhscpg0000388	chr4:41570046-41570447		PHOX2B	TMEM33	0.390
UHNhscpg0000545	chr1:7936516-7937311		TNFRSF9	PARK7	0.165
UHNhscpg0000823	chr3:57516948-57517604	2-PDE	FLJ44290	ARF4	-0.397
UHNhscpg0000843	chr4:153186077-153186612		PET112L	FBXW7	-0.481
UHNhscpg0001058	chr19:42021241-42021797		AK000357	ZNF345	-0.353
UHNhscpg0001183	chr1:93687156-93687645	FNBP1L	DR1	FNBP1L	-0.369
UHNhscpg0001269	chr2:1837452-1837879	MYT1L	BC043230	TSSC1	0.299
UHNhscpg0001470	chr3:142998956-142999372	GRK7	RNF7	ATP1B3	0.276
UHNhscpg0001522	chr15:99009623-99010314		ASB7	LOC440313	0.261
UHNhscpg0002696	chr18:59137439-59137855	BCL2	ZCCHC2	FVT1	0.235
UHNhscpg0003258	chr7:107318613-107319045	DLD	SLC26A3	LAMB1	-0.411
UHNhscpg0003277	chr17:55883441-55883578	APBP2	C17orf84	PPM1D	0.523
UHNhscpg0003348	chr6:13724013-13724309	NOL7	SIRT5	RANBP9	0.825
UHNhscpg0003580	chr17:5220863-5220968	RABEP1	UNQ5783	NUP88	0.213

UHNhscpg0004085	chr15:46725284-46725562	FBN1	CEP152	0.452
UHNhscpg0004100	chr10:119287466-119287753	PDZD8	EMX2	0.250
UHNhscpg0004123	chr22:48741361-48741513	CRELD2	FLJ41993	0.380
UHNhscpg0004255	chr10:124883-124884153	ACADSB	BUB3	-0.525
UHNhscpg0004539	chr9:128414559-128415688	C9orf28	LMX1B	0.369
	chr11:160734105-160734402	SH2D1B	UAP1	-0.528
UHNhscpg0004594	chr11:93333889-93333972	CRSP6	PANX1	0.310
UHNhscpg0004808	chr21:46567171-46567396	C21orf58	PCNT	0.487
UHNhscpg0004936	chr21:31851621-31852410	AB096984	SOD1	0.135
UHNhscpg0005059	chr2:26249810-26250309	RAB10	HADHA	0.241
UHNhscpg0005064				
UHNhscpg0005073	chr4:174679229-174679502	SCRG1	HAND2	0.094
UHNhscpg0005099	chr8:67187866-67188497	DNAJC5B	TRIM55	0.498
UHNhscpg0005199	chr17:8066959-8067843	AURKB	C17orf68	0.411
	chr11:113915762-113915883	REXO2	FAM55D	0.413
UHNhscpg0005710	chr11:73148713-73150271	PLEKHB1	MRPL48	0.400
UHNhscpg0005782	chr14:90930199-90930271	AB040942	KIAA2010	0.369
UHNhscpg0005784	chr8:38648957-38649091	FLJ43582	TACC1	0.360
UHNhscpg0005952	chr12:16561287-16561466	MGST1	LMO3	-0.014
UHNhscpg0005966	chr15:58558595-58558981	NARG2	RORA	0.477
UHNhscpg0005997	chr8:115283071-115283345	CSMD3	TRPS1	0.234
UHNhscpg0006591	chr2:169477583-169478200	G6PC2	ABCB11	0.318
UHNhscpg0007310	chr18:45975419-45975817	AB032945	CCDC11	0.114
UHNhscpg0007521	chr20:34101891-34102561	C20orf152	EPB41L1	0.862
UHNhscpg0007925	chr9:130485904-130486139	WDR34	SET	0.120
UHNhscpg0008102	chr4:184255439-184255717	FAM92A3	WWC2	0.540
UHNhscpg0008621	chr3:83699414-83699648	GBE1	IGSF4D	0.348
UHNhscpg0009095	chrX:48570616-48571239	HDAC6	ERAS	0.496
UHNhscpg0009129	chr19:42434528-42434919	ZNF383	AF289566	0.677
UHNhscpg0009130				

UHNhscpg0009224	chr7:149733492-149733751		MGC33584	GIMAP8	0.199
UHNhscpg0009290	chr13:111768114-111768827		MGC35169	SOX1	0.322
UHNhscpg0009313	chr1:9411460-9411863		SPSB1	MGC4399	0.122
UHNhscpg0009317	chr10:88461721-88462277	LDB3	OPN4	BMPR1A	0.199
UHNhscpg0009329	chr7:135629820-135629963		MTPN	CHRM2	0.072
UHNhscpg0009434	chr15:58851441-58851570	RORA	RORA	VPS13C	0.431
UHNhscpg0009443	chr6:135676890-135676940	AHI1	MYB	AHI1	0.424
UHNhscpg0009445	chr10:99070257-99070534	FRAT1	ARHGAP19	FRAT2	0.748
UHNhscpg0009744	chr7:41865524-41865728		INHBA	GLI3	-0.438
UHNhscpg0009771	chr5:84194219-84194473		EDIL3	BC050328	0.412
UHNhscpg0009839	chr12:93990621-93990967	NR2C1	NDUFA12	FGD6	0.311
UHNhscpg0009846	chr1:89763097-89763481	LRR8B	GBP6	LRR8B	0.176
UHNhscpg0009868	chr4:78960386-78960494		CNOT6L	MRPL1	0.312
UHNhscpg0009900	chr2:66527386-66527591	MEIS1	FLJ16124	ETAA16	0.699
UHNhscpg0010179	chr3:103750158-103750397		ZPLD1	ALCAM	0.388
UHNhscpg0010558	chr2:176585836-176585889		KIAA1715	HOXD13	0.319
UHNhscpg0011073	chr18:72936774-72936872	MBP	MBP	GALR1	0.310
UHNhscpg0011186	chr10:8116616-8116748		ATP5C1	FLJ45983	0.385
UHNhscpg0011502	chr11:59075092-59075158		OR4D11	OSBP	0.423
UHNhscpg0011911	chr1:117711524-117711842	MAN1A2	VTCN1	FAM46C	0.226
UHNhscpg0013020	chr5:37244700-37245302	FLJ13231	FLJ13231	NUP155	0.083
UHNhscpg0013052	chr19:49336853-49337659	ZNF234	ZNF225	ZNF226	0.129
UHNhscpg0013054	chr3:182113379-182113899	FXR1	CCDC39	DNAJC19	0.467
UHNhscpg0013068	chr21:39477469-39478047		DSCR2	BRWD1	0.354
UHNhscpg0013074	chr1:18844025-18844822	PAX7	KLHDC7A	TAS1R2	0.324
UHNhscpg0013083	chr12:46584562-46585348	VDR	HDAC7A	TMEM106C	0.523
UHNhscpg0013316	chr8:59486325-59487010		C8orf72	CYP7A1	0.379
UHNhscpg0013331	chr7:35806170-35806852		FLJ22313	SEPT7	0.314
UHNhscpg0013436	chr3:126257872-126258308		AB033063	SLC12A8	0.317

UHNhscpg0013489	chr5:68498706-68499421	CCNB1	SLC30A5	CENPH	0.108
UHNhscpg0013581	chr5:79366322-79367506	THBS4	MTX3	SERINC5	0.373
UHNhscpg0013708	chr11:62366029-62366350		WDR74	SLC3A2	0.640
UHNhscpg0013713	chr7:135312098-135312915	MTPN	UNQ1940	CHRM2	0.336
UHNhscpg0013792	chr4:165936676-165936857		MARCH1	TRIM61	0.209
UHNhscpg0013793	chr12:53097940-53098512	ITGA5	ZNF385	FAM112B	0.325
UHNhscpg0013871	chr1:61320827-61321489	NFIA	C1orf87	BC034822	0.260
UHNhscpg0013929	chr10:33990718-33991219	NRP1	PARD3		0.085
UHNhscpg0013940	chr21:18539685-18540170	CHODL	C21orf91	PRSS7	0.066
UHNhscpg0013974	chr11:36355255-36356364	FLJ14213	COMMD9	FLJ14213	0.391
UHNhscpg0013977	chr11:458511-459622	PTDSS2	AK128239	AK125720	0.568
UHNhscpg0014018	chr12:105192444-105192779		CKAP4	TCP11L2	0.208
UHNhscpg0014037	chr11:76434479-76435327		B3Gn-T6	CAPN5	0.065
UHNhscpg0014040	chr19:12640945-12641575	PTD008	MAN2B1	MORG1	0.126
UHNhscpg0014051	chr20:49007736-49008634	DPM1	ADNP	MOC53	0.227
UHNhscpg0014056	chr1:35946424-35947119		PSMB2	FLJ38984	0.149
UHNhscpg0014063	chr1:191357430-191358153	CDC73	GLRX2	B3GALT2	0.269
UHNhscpg0014096	chr17:76887872-76888250		MGC15523	C17orf55	0.212
UHNhscpg0014103	chr15:59309160-59309428	RORA	RORA	VPS13C	0.424
UHNhscpg0014115	chrX:33941912-33942633		DMD	TMEM47	0.309
UHNhscpg0014118	chr16:21738689-21739475	RRN3	LOC94431	BC094882	0.180
UHNhscpg0014153	chr11:126090589-126091362	KIRREL3	ST3GAL4	ETS1	0.159
UHNhscpg0014416	chr12:129208994-129209574		AK127723	FZD10	0.327
UHNhscpg0014467	chr2:188864723-188865719	GULP1	TFPI	DIRC1	0.279
UHNhscpg0014470	chr19:6201401-6202238	MLLT1	ACSBG2	ASAH3	0.299
UHNhscpg0014534	chr11:85036396-85036954	TMEM126A	TMEM126B	ZF	0.233
UHNhscpg0014566	chr13:20245319-20245752	LOC221143	IL17D	XPO4	0.460
UHNhscpg0014573	chr18:59138104-59138521		BCL2	FVT1	0.235
UHNhscpg0014617	chr10:90741358-90742025	FAS	STAMBPL1	CH25H	0.291
UHNhscpg0014724	chr11:27340902-27341886	CCDC34	BBOX1	LGR4	0.265

UHNhscpg0014888	chr18:59789147-59789695	SERPINB8	SERPINB2	C18orf20	0.668
UHNhscpg0014945	chr18:11137356-11137477		C18orf58	GNAL	0.248
UHNhscpg0014949	chr10:18980274-18980716	NSUN6	CACNB2	ARL5B	-0.616
	chrX:128867777-				
UHNhscpg0014970	128867977	UTP14A	ZDHHC9	BCORL1	-0.677
UHNhscpg0014994	chr15:99364662-99365508		BC068080	LRRK1	0.310
UHNhscpg0015003	chr19:4382698-4383394	CHAF1A	SH3GL1	UBXD1	0.429
UHNhscpg0015082	chr5:95323531-95324046	ELL2	FIS	PCSK1	0.267
UHNhscpg0015086	chr17:54586478-54587343	FAM33A	TRIM37	PRR11	0.486
UHNhscpg0015097	chr13:20418064-20418641		XPO4	LATS2	0.357
UHNhscpg0015158	chr11:56983677-56984417		SLC43A3	SLC43A1	0.310
	chr1:170016227-				
UHNhscpg0015202	170017078		VAMP4	KIAA0859	0.303
	chr1:144108418-				
UHNhscpg0015290	144108804		NBPF8	HFE2	0.021
UHNhscpg0015303	chr9:137249061-				
UHNhscpg0015475	137249595	ZNF217	AK127261	BC034752	0.232
	chr20:51631246-51632258		AK024093	BC065723	0.244
UHNhscpg0015493	chr12:108822867-	TCHP	GLTP	GIT2	0.258
	108823464				
UHNhscpg0015565	chr5:156501926-	CRSP9	HAVCR2	FAM71B	0.241
UHNhscpg0015865	156502525	PARP8	EMB	ISL1	0.337
UHNhscpg0015967	chr5:4998686-49999611	TGFBR2	RBMS3	GADL1	0.475
UHNhscpg0015971	chr3:30682726-30683315	CBX2	ENPP7	CBX8	0.167
UHNhscpg0016102	chr17:75370070-75370624		FLJ25369	EIF2AK3	0.234
	chr2:88630362-88631132				
UHNhscpg0016107	chr3:143350321-	TFDP2	TFDP2	MGC40579	0.415
	143351137				
UHNhscpg0016151	chr13:112731749-	MCF2L	MCF2L	MCF2L	0.260
UHNhscpg0016153	112732149	ATF1	DIP2B	TMPRSS12	0.336
UHNhscpg0016167	chr12:49444777-49445294	SPATA6	AF251047	AGBL4	-0.044
UHNhscpg0016168	chr1:48556940-48557200		BC050659	PON1	0.459
	chr7:94374935-94375622				
UHNhscpg0016200	chr5:180441377-		BTNL9	TRIM7	0.316
UHNhscpg0016205	180441569		SOX9	SLC39A11	0.286
	chr17:67713392-67713744				
UHNhscpg0016273	chr7:100751449-	RABL5	FIS1	AK124120	0.423
UHNhscpg0016274	100752104	MYCBP	RRAGC	GJA10	0.544
UHNhscpg0016365	chr1:39110922-39111736		REXO1L1	ATP6V0D2	0.145
UHNhscpg0016389	chr8:87030008-87030494	VKORC1	ZNF646	CR622376	0.591

UHNhscpg0016511	chr17:30594652-30595203	MGC19764	AMAC1	FLJ34922	0.358
UHNhscpg0016521	chr3:139635428-139635930		MRAS	FAM62C	0.252
UHNhscpg0016617	chr7:20604954-20605616		ITGB8	AY851364	0.664
UHNhscpg0016690	chr11:128747666-128747940		AK127847	BARX2	0.220
UHNhscpg0016735	chr1:68057841-68058291	SAPS3	LRP5	GAL	0.602
UHNhscpg0016749	chr8:123801107-123801380		HAS2	ZHX2	0.325
UHNhscpg0016866	chr4:129950729-129951004	PHF17	PGRMC2	PHF17	0.646
UHNhscpg0016867	chr4:76078566-76079122	DKFZP564O0823	BTC	RCHY1	0.432
UHNhscpg0016870	chr5:54215029-54215299		SNAG1	ESM1	0.467
UHNhscpg0016961	chr2:239988598-239988994		HDAC4	FLJ45964	0.384
UHNhscpg0017300	chr22:43987533-43988137	C22orf9	C22orf9	BC108900	0.405
UHNhscpg0017328	chr8:99126095-99126507	RPL30	MATN2	C8orf47	0.470
UHNhscpg0017558	chr3:131175719-131175943		TMCC1	TRH	0.541
UHNhscpg0017564	chr12:14847117-14848131	WBP11	H2AFJ	ART4	0.476
UHNhscpg0017583	chr4:81770204-81770448	AK128642	FGF5	BMP3	0.421
UHNhscpg0017604	chr21:39607959-39608495	AF426265	BRWD1	HMGN1	0.480
UHNhscpg0017657	chr13:19815231-19815569		GJB6	BC071810	0.938
UHNhscpg0017711	chr1:146203056-146203351	BC034418	AK128227	FLJ39739	0.843
UHNhscpg0017725	chr3:1808180805651	NDUFB5	MRPL47	USP13	0.264
UHNhscpg0017768	chr3:101721999-101722586	TMEM45A	TOMM70A	GPR128	0.323
UHNhscpg0017793	chr16:84162460-84162782		MGC22001	KIAA0182	0.192
UHNhscpg0017855	chr4:26695450-26695704		STIM2	FLJ45721	0.351
UHNhscpg0018022	chr6:7072268-7072514	RREB1	AK127740	RREB1	0.178
UHNhscpg0018024	chr12:79626214-79626959	MYF6	FLJ90579	MYF5	-0.371
UHNhscpg0018048	chr17:61122273-61122774	CCDC46	AXIN2	APOH	0.653
UHNhscpg0018198	chr1:15725895-15726441	DNAJC16	CASP9	AGMAT	0.147
UHNhscpg0018227	chr2:135526438-135526876	RAB3GAP1	YSK4	ZRANB3	0.210
UHNhscpg0018235	chr14:88753793-88754227	CHES1	CHES1	AK130927	0.459
UHNhscpg0018310	chr6:8280993-8282026		AF217970	THEX1	0.122
UHNhscpg0018426	chr3:10755074-10755751		ATP2B2	SLC6A11	0.217
UHNhscpg0018454	chr17:60262879-60263369	AL833315	AK123704	FLJ34306	0.393

UHNhscpg0018616	chr11:9642277-9642997	SWAP70	WEE1	AK125613	0.509
UHNhscpg0018650	chr6:10727741-10728273	GCNT2	MGC40222	AK096781	-0.008
UHNhscpg0018747	chr16:84146341-84146647		MGC22001	KIAA0182	0.214
UHNhscpg0018781	chr22:36906980-36907601	PLA2G6	BAIAP2L2	MAFF	-0.040
UHNhscpg0018805	chr19:53640757-53641312	GRWD1	GRIN2D	KCNJ14	0.306
UHNhscpg0018872	chr22:28426950-28427806		NF2	CABP7	0.319
UHNhscpg0018886	chr18:51598971-51599075	AK127787	TCF4	AK127645	0.129
	chr5:114368189-114368399		KCNN2	TRIM36	0.528
UHNhscpg0018912	chr19:20995208-20995601	ZNF430	ZNF85	ZNF714	0.139
UHNhscpg0018929	chr1:48332408-48332706		FOXD2	AF111848	0.301
UHNhscpg0018985	chr5:4348162-4348505		IRX1	ADAMTS16	0.208
UHNhscpg0019001	chr4:158361558-158361804	GRIA2	GLRB	C4orf18	0.355
UHNhscpg0019014	chr2:54536224-54536496		FLJ40298	SPTBN1	0.482
UHNhscpg0019016	chr20:33792864-33793374	RNPC2	C20orf52	PHF20	0.106
UHNhscpg0019117	chr17:24079651-24080467	NEK8	TLCD1	TRAF4	0.120
UHNhscpg0019151	chr2:177647276-177647897		MTX2	HNRP3	0.331
UHNhscpg0019319	chr19:40227874-40229005	HPN	SCN1B	AK124779	0.262
UHNhscpg0019417	chr8:38081162-38081389		EIF4EBP1	ASH2L	0.133
UHNhscpg0019472	chr11:127646621-127647101		KIRREL3	ETS1	0.177
UHNhscpg0019513	chr10:1444144463	ADARB2	AK125752	AK129565	0.196
UHNhscpg0019539	chr8:133999422-133999655	TG	PHF20L1	SLA	0.435
UHNhscpg0019553	chr6:33655547-33656258	BAK1	ZBTB9	FLJ43752	0.227
UHNhscpg0019676	chr2:176723074-176723546		HOXD8	HOXD4	0.083
UHNhscpg0019681	chr7:36661768-36662660	AOAH	ANLN	ELMO1	0.382
UHNhscpg0019748	chr19:32423794-32424129		ZNF539	BC068609	0.294
UHNhscpg0019836	chr15:33316668-33317005		AK094481	AK128197	0.281
UHNhscpg0020198	chr12:106603370-106604499	PWP1	BTBD11	PRDM4	0.241
UHNhscpg0020329	chr12:54925037-54925570	AK091555	AY358160	COQ10A	0.098
UHNhscpg0020372	chr8:91729041-91729327		TMEM64	EFCBP1	0.292
UHNhscpg0020409	chr8:142314675-142315035		SLC45A4	GPR20	0.222
UHNhscpg0020507	chr18:12236532-12237082		IMPA2	CIDEA	0.504
UHNhscpg0020852	chr9:95070562-95071146	WNK2	NINJ1	AK124314	0.231
UHNhscpg0020967					

Table 3

Overlap of 50 dysregulated genes identified from the CpG island microarray analysis and global gene expression profiling previously performed using the same samples. Log2 ratios of expression values were obtained from gene expression data (GEO accession # GSE4187) and averaged across all replicates corresponding to the same samples used for CpG island microarray analysis. CpG methylation log2 ratios were calculated as described in Methods. Genes both upstream and downstream of the CpG island were included in the overlap analysis, in addition to genes that had direct overlap with the differentially methylated CpG island.

R96310	2'-PDE	2'-phosphodiesterase	+	0	-0.397	-0.131
T71316	ARF4	ADP-ribosylation factor 4	-	14525	-0.416	-0.207
AA425722	ARHGAP19	Rho GTPase activating protein 19	-	27854	0.748	-0.130
H43438	ARHGAP19	RHO GTPASE ACTIVATING PROTEIN 19	-	27854	0.748	-0.153
W63749	BCL2	B-cell CLL/lymphoma 2	-	511	0.235	-0.286
N27448	BMPRI1A	BONE MORPHOGENETIC PROTEIN RECEPTOR, TYPE IA	+	44098	0.234	0.405
AA670419	C22orf9	chromosome 22 open reading frame 9	-	0	0.405	-0.160
AA670419	C22orf9	chromosome 22 open reading frame 9	-	522	0.405	-0.160
W20354	CRSP6	cofactor required for Sp1 transcriptional activation, subunit 6, 77kDa	+	147787	0.310	-0.154
AA918102	DPM1	dolichyl-phosphate mannosyltransferase polypeptide 1, catalytic subunit	-	0	0.227	-0.144
AA252470	DR1	down-regulator of transcription 1, TBP-binding (negative cofactor 2)	+	86422	-0.369	0.201
AA436178	EIF2AK3	eukaryotic translation initiation factor 2-alpha kinase 3	-	6243	0.249	-0.062
AA418818	ELL2	elongation factor, RNA polymerase II, 2	-	0	0.267	-0.245
T87150	ELL2	ELONGATION FACTOR, RNA POLYMERASE II, 2	-	0	0.267	-0.348
AA058597	FAM46C	family with sequence similarity 46, member C	+	238284	0.157	-0.924
W46575	FGF5	FIBROBLAST GROWTH FACTOR 5	+	339010	0.421	-0.157
H80525	FLJ34922	hypothetical protein FLJ34922	-	106241	0.281	-0.210
AA430662	FLJ38984	hypothetical protein FLJ38984	-	4944	0.207	-0.284
AA927951	FXR1	fragile X mental retardation, autosomal homolog 1	+	0	0.467	-0.238
AI623173	GAL	GALANIN	+	150319	0.546	-0.092
AA496097	HNRPA3	glucan (1,4-alpha-), branching enzyme 1 (glycogen branching enzyme, Andersen R09069 GBE1 disease, glycogen storage disease type IV)	-	1805979	0.348	-0.309
N62301	IGSF4D	heterogeneous nuclear ribonucleoprotein A3	+	137848	0.346	-0.210
AA101146	KIAA1715	immunoglobulin superfamily, member 4D	-	1391174	0.383	-0.199
AA431967	LATS2	KIAA1715	-	10572	0.319	-0.122
		LATS, large tumor suppressor, homolog 2 (Drosophila)	-	27937	0.334	0.105
		low density lipoprotein receptor-related R83038 protein 5	+	84524	0.602	-0.173
	LRP5		+	14686	0.262	0.247
AI240161	LRRK1	leucine-rich repeat kinase 1	+	14686	0.262	0.149
R51605	LRRK1	leucine-rich repeat kinase 1	+	0	0.226	-0.098
AA455062	MAN1A2	mannosidase, alpha, class 1A, member 2	+	0	0.226	-0.165
R22905	MAN1A2	mannosidase, alpha, class 1A, member 2	-	412024	0.209	-0.089
AI077321	MARCH1	membrane-associated ring finger (C3HC4) 1	-	4181	0.212	-0.149
R46332	MGC15523	HYPOTHETICAL PROTEIN MGC15523				

AA495936	MGST1	microsomal glutathione S-transferase 1	+	152677	-0.014	0.426
AA626777	MRAS	muscle RAS oncogene homolog	+	28363	0.252	-0.188
AA099873	MYCBP	c-myc binding protein	-	0	0.544	-0.183
AA098867	NRP1	neuropilin 1	-	326879	0.085	-0.185
AA464606	PANX1	pannexin 1	+	167769	0.365	-0.151
AA047567	PGRMC2	progesterone receptor membrane component 2	-	522022	0.646	-0.133
AA136664	PHF17	PHD finger protein 17	+	926	0.671	-0.105
AA505136	PHF17	PHD finger protein 17	+	926	0.671	-0.116
AA136664	PHF17	PHD finger protein 17	+	0	0.646	-0.105
AA505136	PHF17	PHD finger protein 17	+	0	0.646	-0.116
AI015442	PHF20L1	PHD finger protein 20-like 1	+	70138	0.435	-0.131
PROTEASOME (PROSOME, MACROPAIN)						
N54157	PSMB2	SUBUNIT, BETA TYPE, 2	-	66694	0.149	-0.155
AA485992	PWP1	nuclear phosphoprotein similar to S. cerevisiae PWP1	+	0	0.241	-0.116
H95393	RAB6A	RAB6A, member RAS oncogene family	-	0	0.400	-0.138
N71050	RORA	RAR-RELATED ORPHAN RECEPTOR A	-	17774	0.572	-0.181
N71050	RORA	RAR-RELATED ORPHAN RECEPTOR A	-	0	0.120	-0.181
N71050	RORA	RAR-RELATED ORPHAN RECEPTOR A	-	144417	0.120	-0.181
AA485141	SLA	Src-like-adaptor	-	118500	0.451	-0.094
AA404709	SLC43A3	solute carrier family 43, member 3	-	32578	0.310	-0.131
AA453898	ST3GAL4	ST3 beta-galactoside alpha-2,3-sialyltransferase 4	+	274627	0.159	-2.157
AA432278	THBS4	thrombospondin 4	+	0	0.373	-0.103
AA676378	TMCC1	transmembrane and coiled-coil domains 1	-	93720	0.541	0.120
N68465	UAP1	UDP-N-acetylglucosamine pyrophosphorylase 1	+	63544	-0.471	-0.234
N51629	VAMP4	vesicle-associated membrane protein 4	-	38390	0.303	-0.090
AA485226	VDR	vitamin D (1,25-dihydroxyvitamin D3) receptor	-	0	0.523	-0.081
AA625788	VKORC1	vitamin K epoxide reductase complex, subunit 1	-	0	0.591	-0.094
AA130669	WBP11	WW domain binding protein 11	-	0	0.476	-0.122
AA678088	XPO4	exportin 4	-	5262	0.495	-0.077
AA678088	XPO4	exportin 4	-	43180	0.357	-0.077

Table 4 Primer sequences

Gene			
Symbo	Application	Forward Primer (5' --> 3')	Reverse Primer (5' --> 3')
BCL2	qPCR	TATTGTGGCTGCACTTGCTC	TGTTGCCCAACTGCAAAATA
RORA	qPCR	ACACCAGCATCAGGCTTCTT	GGTCTGCCACGTTATCTGCT
	Bisulfite sequencing		
BCL2	Nested PCR- Outer	GGTGGTTTAGAGGAGGGTTTTT	ACAAAAATCCTCTTCTAATTAAACTC
	Bisulfite sequencing		
BCL2	Nested PCR- Inner	GAGAATGAAGTAAGAGGATAGGT	CACCCTTTCTCCTCCTCCTAATC
RORA	MSP: Methylated	GGTTGGAGAAGTTTTTCGTTAGC	GACGAACGAACAAACAAAACG
	MSP: Unmethylated	TTGGTTGGAGAAGTTTTTGTTAGT	CAAACAAAACACAAAAAACACA
Forward Primers SEQ ID NOS: from top to bottom: 3, 5, 7, 9, 11, 13			
Reverse Primers SEQ ID NOS: from top to bottom: 4, 6, 8, 10, 12, 14			

Table 5: Four high level functions associated with genes differentially methylated between discordant monozygotic twins and identified by CpG island microarray. P values represent significance by Fisher Exact Test and indicates the probability that the indicated function is not associated with the dataset of differentially methylated genes (Ingenuity Pathway Analysis software).

Category (# genes)	Function Annotation	Molecules	p-value
Neurological Disease (Lopez-Rangel and Lewis, (2006) <i>Clin Genet</i> 69, 21-2)	Huntington's disease	ELMO1, EMX2, FOXN3, GABRA4, MEIS2, NDUFB5, XPA	2.62E-04
	neurodegeneration of CA3 neurons	GRIA2	2.23E-03
	congenital hydrocephalus of mice	E2F5	4.46E-03
		E2F5, ELMO1, EMX2, FOXN3, GABRA4, GRIA2, MEIS2, NDUFB5, TIAM1, TOP2A, XPA	6.24E-03
	neurological disorder	XPA	8.91E-03
	hypoplasia of cerebellum	E2F5, EMX2	1.61E-02
	Encephalopathy	GABRA4, GRIA2	2.01E-02
	major depression	GABRA4, GRIA2	2.09E-02
	epilepsy, seizures		
	cell death of pyramidal neurons	GRIA2	2.87E-02
	Dyskinesia	GRIA2	2.87E-02
	progressive motor neuropathy	GABRA4, GRIA2, TOP2A	2.95E-02
Nervous System Development and Function (Beaudet, A. L. (2007). <i>Nature Medicine</i> 13, 534-536)	calcium permeability of neurons, innervation of Purkinje cells	GRIA2	2.23E-03

	secretion of cerebrospinal fluid	E2F5	2.23E-03
	dedepression and depotentiation of CA1 neuron	GRIA2	4.46E-03
	formation of cerebellar folia,		
	Size of dendrites, neurogenesis	XPA	4.46E-03
	size of hippocampus	EMX2	6.69E-03
	formation of neurite branches	TIAM1	8.91E-03
	size of cerebral cortex	EMX2	8.91E-03
	activation of synapse, LTD of CA1 neurons, Purkinje cells, and mossy fibers	GRIA2	1.33E-02
	development of dentate gyrus	EMX2	11.77E-02
	synaptic transmission of pyramidal neurons	GRIA2	1.77E-02
	Memory	GRIA2, ITGA5	2.62E-02
	migration of cortical neurons and Schwann cells	TIAM1	3.52E-02
	morphology of neurons	TIAM1	4.38E-02
Cellular Assembly and Organization (Schanen, (2006) Hum Mol Genet 15 Spec No 2,	retraction of glial cell projections	GRIA2	2.23E-03
	structural integrity of nucleosomes	TOP2A	4.46E-03
	retraction of plasma membrane projections	GRIA2, TIAM1	8.93E-03
	size of growth cone	TIAM1	1.99E-02
	formation of actin stress fibers	ITGA5, TIAM1	4.83E-02
Embryonic Development (Folstein and Rutter, (1977) J Child Psychol Psychiatry 18, 297-321)	degeneration of notochord	ITGA5	2.23E-03
	formation of gonadal ridge	CBX2	2.23E-03
	neurogenesis of embryonic tissue	XPA	6.69E-03
	cell death of neural crest cells	ITGA5	1.77E-02
	cleavage of embryo	TOP2A	1.77E-02

Table 6: Four high level functions associated with genes differentially methylated between monozygotic twins and unaffected siblings and identified by Cpg island microarray. P values represent significance by Fisher Exact Test and indicates the probability that the indicated function is not associated with the dataset of differentially methylated genes (Ingenuity Pathway Analysis software).

Category (# genes)	Function Annotation	Molecules	p-value
Gene Expression (Nagarajan et al. (2006) Epigenetics 1, 172-182)	Transcription	ATF1, CBX2, CCNB1, CHAF1A, FOXN3, FRAT1, ILK, IRF1, ITGA5, MED7 (includes EG:9443), MEIS1, MLLT1, MTPN, MYF6, NFIA, NR2C1, PAX7, PWP1, RBM39, RORA, RREB1, RRN3, SET, TFDP2, TG, TGFB2, VDR, ZNF234	9.02E-08
	Transactivation	ATF1, FRAT1, IRF1, NFIA, PAX7, RBM39, RORA, TFDP2, TGFB2, VDR	1.92E-03
	activation of TGF beta response element	ITGA5, TGFB2	4.93E-03
	transactivation of response element	VDR	5.29E-03
	transactivation of RORE binding site	RORA	5.29E-03
	expression of DNA endogenous promoter	TG, VDR	7.06E-03
	activation of Atf-1 binding site	ATF1	1.05E-02
	activation of NF1 binding site	NFIA	1.05E-02
	binding of TREpal motif	VDR	1.05E-02
	transcription of beta catenin response element	FRAT1	1.05E-02
	activation of HAF1/HAF1a binding site	IRF1	1.58E-02
	activation of IRF1/IRF2 binding site	IRF1	1.58E-02
	binding of NFAT response element	VDR	1.58E-02
Cell Death (Klose and Bird (2006) Trends in Biochemical Sciences 31, 89- 97)	Apoptosis	ATF1, BAK1, CCNB1, CDC73, FAS, FRAT1, FXR1, GRIA2, GULP1, HPN, ILK, IRF1, ITGA5, MCF2L, MED7, MTPN, PAX7, PLA2G6, RRN3, TGFB2, TIAM1, VDR	4.30E-03
Cell Death	delay in cell death of cortical neurons	BAK1	5.29E-03
Cell Death	killing of oligodendrocytes	MBP	5.29E-03
Cell Death	cell death of microglia	FAS, IRF1	5.43E-03

Cell Death	apoptosis of embryonic cells	ATF1, FAS, ITGA5 ATF1, BAK1, CCNB1, CDC73, FAS, FRAT1, FXR1, GRIA2, GULP1, HPN, ILK, IRF1, ITGA5, MBP, MCF2L, MED7, MEIS1, MTPN, PAX7, PLA2G6, RRN3	6.02E-03
Cell Death	cell death	TGFBR2, TIAM1, VDR	2.56E-02
Cell Death	apoptosis of embryonic stem cells	FAS, ITGA5	6.61E-03
Cell Death	cell death of brain cells	BAK1, FAS, FRAT1, GRIA2	7.65E-03
Cell Death	cell death of cortical neurons	BAK1, FAS, GRIA2	2.54E-02
Neurological Disease (Kaminsky et al. (2006) Ann Med 38, 530-44)	delay in cell death of cortical neurons	BAK1	5.29E-03
	killing of oligodendrocytes	MBP	5.29E-03
	neurodegeneration of CA3 neurons	GRIA2	5.29E-03
	cell death of microglia	FAS, IRF1	5.43E-03
	disease of brain cells	GRIA2, TG	1.02E-02
	depletion of neurons	DLD	1.05E-02
	gracile axonal dystrophy	PLA2G6	1.05E-02
	mobility of brain cancer cell lines	ITGA5	1.05E-02
	multiple sclerosis of humans	MBP	1.58E-02
	Neurodegeneration	GRIA2, PLA2G6, TGFBR2	1.76E-02
	apoptosis of granule cells	FAS, FRAT1	1.89E-02
	cell death of brain cells, cortical neurons	BAK1, FAS, FRAT1, GRIA2	2.54E-02
	apoptosis of neural precursor cells	FAS	2.62E-02
	hypomyelination of axons	MBP	2.62E-02
Nervous System Development and Function (Beaudet, A. L. (2007) Nature Medicine 13, 534-536)	calcium permeability of neurons, innervation of Purkinje cells, retraction of glial cell projections	GRIA2	5.29E-03
	formation of hippocampal commissure, presence/development of corpus callosum, formation of sling cells	NFIA	5.29E-03
	dedepression and depotentiation of CA1 neurons	GRIA2	1.05E-02
	organization of axons, organization of central nervous system	MBP	1.05E-02
	quantity of satellite cells	PAX7	1.05E-02
	dilation of cerebral ventricles	NFIA	1.58E-02
	function of brain	RORA	2.10E-02
	formation of neurite branches	TIAM1	2.10E-02
	migration of neuroglia	NFIA, TIAM1	2.26E-02

WE CLAIM:

1. An array comprising a plurality of different oligonucleotides with specificity for at least one of the genes contained in Table 1, Table 2, or Table 3, or a combination thereof.
2. The array of claim 1, wherein members of said plurality have specificity for at least one of the genes selected from retinoic acid receptor related orphan receptor (RORA) and B-cell CLL/Lymphoma 2 (BCL-2) genes.
3. The array of claims 1 or 2, comprising oligonucleotides having specificity for differentially methylated regions of the genes or for differentially methylated CpG islands of promoter(s) of the gene(s).
4. A method of screening a subject for a neurological disease or disorder comprising the steps of:
 - (a) isolating nucleic acid from a subject;
 - (b) measuring a level of DNA methylation of at least one gene in Table 1, Table 2, Table 3 or a combination thereof in the isolated nucleic acid of (a); and
 - (c) diagnosing the subject as being at risk for or affected by a neurological disease or disorder if there is a statistically significant difference in the level of DNA methylation of at least one gene listed in Table 1, Table 2, Table 3 of (a) compared to the level of DNA methylation of the same genes from a sample from a healthy individual not having the neurological disease or disorder.
5. The method of Claim 4, wherein the neurological disease is an autism spectrum disorder.
6. The method of claim 5 wherein the autistic spectrum disorder is selected from the group consisting of pervasive developmental disorder-not otherwise specified (PDD-NOS), atypical autism, and Asperger's Disorder.
7. The method of Claim 4, wherein the at least one gene is selected from genes involved in nervous system development, axon guidance, synaptic transmission or plasticity, myelination, long-term potentiation, neuron toxicity, Purkinje cell differentiation, cerebellar development, embryonic development, regulation of actin networks, digestion, inflammation, oxidative stress, epilepsy, apoptosis, morphogenesis, cell survival, differentiation, the unfolded protein response, Type II diabetes and insulin signaling, digestion, liver toxicity (hepatic stellate cell activation, fibrosis, and cholestasis), endocrine function, circadian rhythm, cholesterol metabolism and the steroidogenesis pathway, or a combination thereof.
8. The method of Claim 4, wherein the healthy individual is a non-phenotypic discordant twin, sibling of the subject, or unrelated subject.

9. The method of Claim 4, wherein the neurological disease or disorder is an autism spectrum disorder characterized by (a) a lower severity score across all ADIR items, (b) an intermediate severity score across all ADIR items, (c) a higher severity scores on spoken language items on the ADIR, (d) a higher frequency of savant skills, (e) a severe language impairment, or a combination thereof.

10. The method of Claim 4, wherein DNA methylation is quantified with an assay comprising (a) bisulfite-sequencing, (b) methylation-specific PCR (MSP), (c) large scale CpG island microarray analysis using methylation-enriched and -unenriched samples, (d) promoter analysis using methylation-enriched and -unenriched samples, (e) combined bisulfite and restriction analysis (COBRA), (f) pyrosequencing, (g) Methyl-Light (Premier Biosoft International) or Methyl-Profiler (SABiosciences) methylation analysis, (h) Sequenom mass analyzer, (i) bisulfite treatment of DNA combined with selected amplification of the specific promoter regions of differentially methylated genes with the addition of a T7 promoter tag, followed by T7-mediated transcription and RNase T1 cleavage of the transcript with methylation sites determined by MALDI-TOF mass spectrometry analysis in, or a combination thereof.

11. A method for determining a differential DNA methylation profile for at least one autism spectrum disorder, comprising

- (a) obtaining samples of control and experimental DNA, wherein the experimental DNA is generated from a sample isolated from a subject afflicted with or suspected of being afflicted with the at least one autism spectrum disorder and the control DNA is generated from a sample isolated from a healthy individual;
- (b) applying the samples of (a) to a microassay comprising a plurality of different oligonucleotides having specificity for differentially-methylated genes associated with the at least one autism spectrum disorder to allow hybridization between the oligonucleotides and the control DNA and experimental DNA; and
- (c) identifying the oligonucleotides on the microarray that display differential hybridization to the experimental DNA relative to the control DNA, thereby determining a differential DNA methylation profile for the at least one autism spectrum disorder.

12. The method according to Claim 11, wherein the plurality of different oligonucleotides is specific for at least one differentially DNA-methylated gene in Table 1, Table 2, Table 3 or a combination thereof.

13. The method of Claim 12, wherein the at least one autism spectrum disorder is selected from autistic disorder, pervasive developmental disorder-not otherwise specified (PDD-NOS), atypical autism, and Asperger's Disorder.

14. A method for distinguishing between different phenotypes of an autism spectrum disorder comprising severely language impaired (L), mildly affected (M), or "savants" (S) comprising

- (a) obtaining samples of control and experimental DNA, wherein the experimental DNA is generated from a sample isolated from a subject suspected of being afflicted with at least one phenotype selected from the severely language impaired (L), mildly affected (M), or "savants" (S);
- (b) applying the samples of (a) to a microarray comprising a plurality of different oligonucleotides having specificity for differentially DNA-methylated genes associated with the at least one phenotype to allow hybridization between the oligonucleotides and the control and experimental DNA; and
- (c) identifying the oligonucleotides on the microarray which display differential hybridization to the experimental DNA relative to the control DNA thereby determining a DNA methylation profile for distinguishing among the different phenotypes of autism spectrum disorder.

15. The method according to Claim 14, wherein the plurality of different oligonucleotides is specific for at least one differentially methylated gene in Table 1, Table 2, Table 3 or a combination thereof.

16. The method of Claim 14, wherein the at least one autism spectrum disorder is selected from autistic disorder, pervasive developmental disorder-not otherwise specified (PDD-NOS), atypical autism, or Asperger's Disorder.

17. The method of claims 14 wherein the control DNA is from a healthy individual not having an autism spectrum disorder and wherein the difference in the DNA methylation profiles distinguishes the different phenotypes of autism spectrum disorder.

18. A method of assessing the efficacy of a treatment in a subject having at least one autism spectrum disorder comprising,

- (a) determining a level of DNA methylation of at least one gene in Table 1, Table 2, Table 3 or a combination thereof in a sample of a test subject, or in a plurality of samples of a selected tissue type of a test subject to generate a methylation profile for the sample;
- (b) determining a degree of similarity between (i) the level of DNA methylation of the profile of (a); and (ii) the level of DNA methylation of at least one differentially DNA-methylated gene in Table 1, Table 2, Table 3 or a combination thereof in a control profile,

wherein the control profile is a methylation profile specific for the genes set out in listed in Table 1, Table 2, Table 3, or a combination thereof, produced by a therapy which has been

shown to be efficacious in treatment of the at least one autism spectrum disorder; wherein a high degree of similarity of the differential methylation profile data is indicative that the treatment is effective.

19. The method of claim 18 wherein the high degree of similarity is log 2 ratio less than +/- 0.3.

20. A method of assessing the efficacy of a treatment in a subject having at least one autism spectrum disorder comprising

- (a) determining DNA methylation profile data specific for at least one differentially DNA-methylated gene contained in Table 1, Table 2, Table 3 or a combination thereof, in a plurality of subject samples of a selected tissue type, wherein the subject has received a treatment, by measuring the level of DNA methylation of the at least one differentially DNA-methylated gene;
- (b) determining a degree of similarity between (i) the DNA methylation profile data in the subject samples of (a); and (ii) DNA methylation profile data specific for the genes contained in Table 1, Table 2, Table 3 or a combination thereof, produced by a therapy which has been shown to be efficacious in treatment of the at least one autism spectrum disorder; wherein a high degree of similarity of the DNA methylation profile data of (i) and (ii) is indicative that the treatment is effective.

21. The method of claim 20 wherein the high degree of similarity is a log2 ratio less than +/- 0.3.

22. A method of identifying a DNA methylation profile responsive to administration of a therapeutic treatment to a subject with at least one autism spectrum disorder comprising,

- (a) obtaining samples of control and experimental nucleic acid, wherein the experimental nucleic acid is generated from a sample isolated from a test subject (i) prior to receiving a therapeutic treatment and (ii) after receiving the therapeutic treatment;
- (b) preparing one or more microarrays comprising a plurality of different oligonucleotides, wherein the oligonucleotides are specific for genes associated with an autism spectrum disorder;
- (c) applying the prepared samples to the one or more microarrays to allow hybridization between the oligonucleotides and the control and experimental nucleic acids;
- (d) identifying the oligonucleotides on the microarray that display differential hybridization to the experimental nucleic acid relative to the control nucleic acid to generate a differential DNA methylation profile for the samples, and
- (e) comparing the profiles of the samples,

wherein increase in the similarity of the differential DNA-methylation of the test sample of (ii) and the control as compared to the similarity of the differential DNA-methylation profile of (i) and control identifies a differential DNA methylation profile responsive to the administration of the therapeutic treatment to the subject with at least one autism spectrum disorder.

23. The method of Claim 22 wherein the control sample is taken from the subject prior to the therapeutic treatment or from a healthy subject not having an autism spectrum disorder or a subject have a less severe form of autism spectrum disorder.

24. The method of Claim 21, wherein the plurality of different oligonucleotides is specific for at least one differentially DNA-methylated gene set out in Table 1, Table 2, Table 3 or a combination thereof.

25. The method of Claim 22, wherein the at least one autism spectrum disorder neurological condition comprises autistic disorder, pervasive developmental disorder-not otherwise specified (PDD-NOS), atypical autism, Asperger's Disorder, or a combination thereof.

26. A method for conducting drug discovery comprising,

- (a) generating a database of DNA methylation profile data representative of the genetic expression response of at least one selected neuronal tissue type from a subject that was subjected to at least one of a plurality of behavioral therapies and that has undergone a selected physiological change since commencement of the behavioral therapy;
- (b) administering a small molecule test agent(s) to untreated subjects to obtain DNA methylation profile data associated with administration of the agent(s) and comparing the obtained data with the one or more DNA methylation profiles of (a);
- (c) selecting the test agent(s) that induce DNA methylation profiles similar to the DNA methylation profiles obtained by administration of the behavioral therapy;
- (d) conducting therapeutic profiling of the selected test agent(s), or analogs thereof, for efficacy and toxicity in subjects; and
- (e) identifying a pharmaceutical preparation including one or more agent(s) identified in step (d) as having an acceptable therapeutic and/or toxicity profile.

27. The method of Claim 26, wherein the behavioral therapy comprises applied behavior analysis (ABA) intervention methods, dietary changes, exercise, massage therapy, group therapy, talk therapy, play therapy, conditioning, or alternative therapies such as sensory integration and auditory integration therapies.

28. The method of Claim 26, wherein the selected physiological change includes one or more improvements in social interaction, language abilities, restricted interests, repetitive behaviors, sleep

disorders, seizures, gastrointestinal, hepatic, and mitochondrial function, neural inflammation, or a combination thereof.

29. The method of Claim 26, wherein, prior to administration of behavioral therapy, the subject showed at least one symptom of a psychological or physiological abnormality.

30. The method of Claim 26, wherein the neuronal tissue type is selected from the group consisting of olfactory bulb cells, cerebrospinal fluid, hypothalamus, amygdala, pituitary, nervous system, brainstem, cerebellum, cortex, frontal cortex, hippocampus, striatum, and thalamus

31. The method of any one of Claims 4, 11, 14, 18, 20, 23, or 26, wherein the gene comprises any one or more of *RORA* or *BCL-2*.

32. The method of any one of Claims 4, 11, 14, 18, 20, 23, or 26, wherein the sample is a blood sample.

33. The method of any one of Claims 4, 11, 14, 18, 20, 23, or 26, wherein the sample comprises lymphocytes.

34. The method of any one of Claims 4, 11, 14, 18, 20, 23, or 26, wherein the sample is an LCL derived from the subject.

35. A kit for identifying a compound for treating at least one autism spectrum disorder comprising (a) a database having information stored therein of one or more DNA methylation profiles specific for the genes set out in Table 1, Table 2, Table 3 or a combination thereof, of subjects that have been subjected to at least one of a plurality of selected autism spectrum disorder neurological therapies and wherein the subject has undergone a desired physiological change; and (b) a computer program for comparing DNA methylation profile data obtained from assays wherein a test compound is administered to a subject with the database of stored DNA methylation profiles and providing information representative of a measure of similarity between the DNA methylation profile data and one or more stored DNA methylation profiles.

36. A computer-readable medium on which is encoded programming code for analyzing autism spectrum disorder from a plurality of data points comprising a profile of differentially expressed DNA methylation genes, wherein said differential DNA methylation profile is specific for at least one differentially DNA methylated gene set out in Table 1, Table 2, Table 3, or a combination thereof.

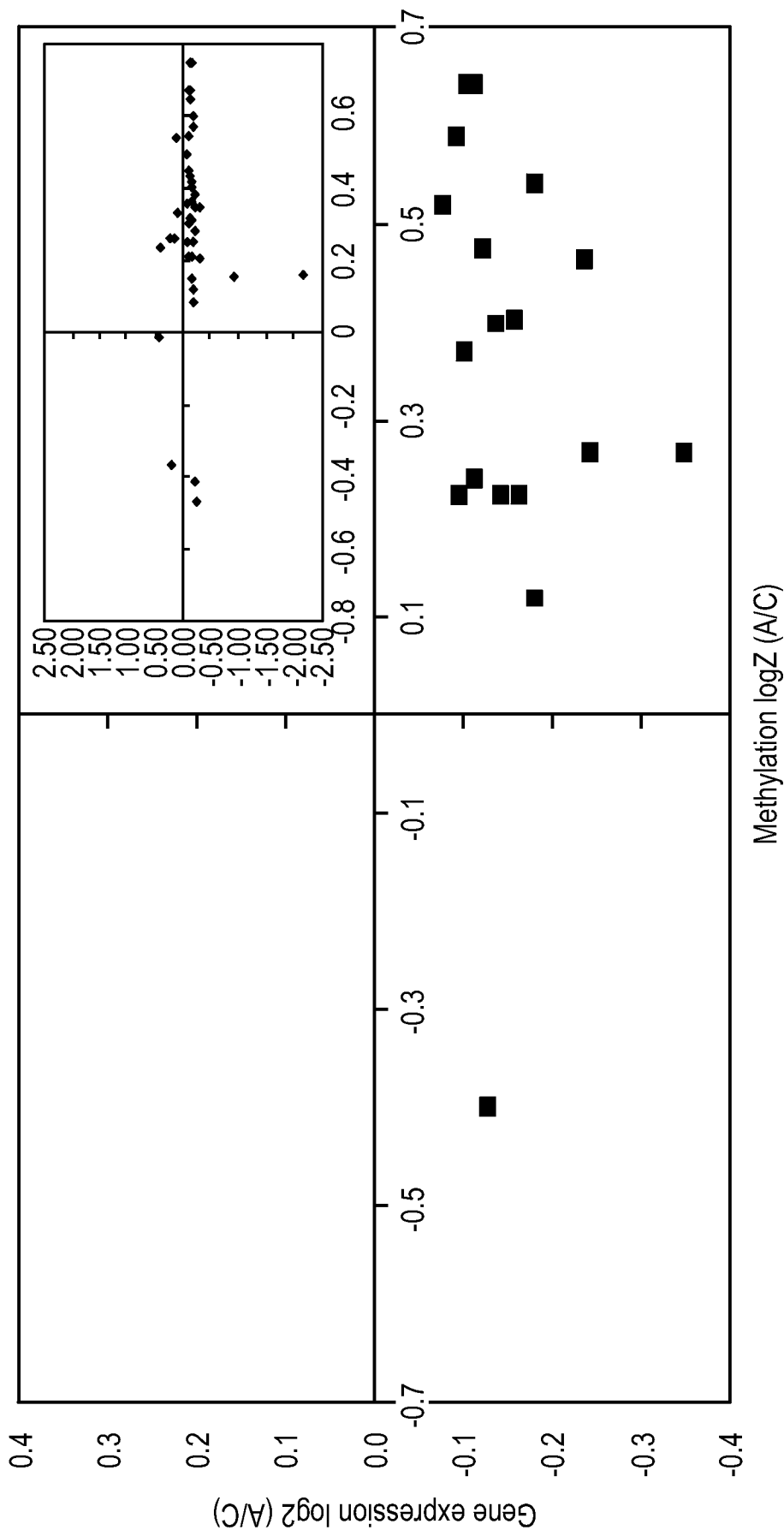


FIG. 1A

2/8

FIG. 1B

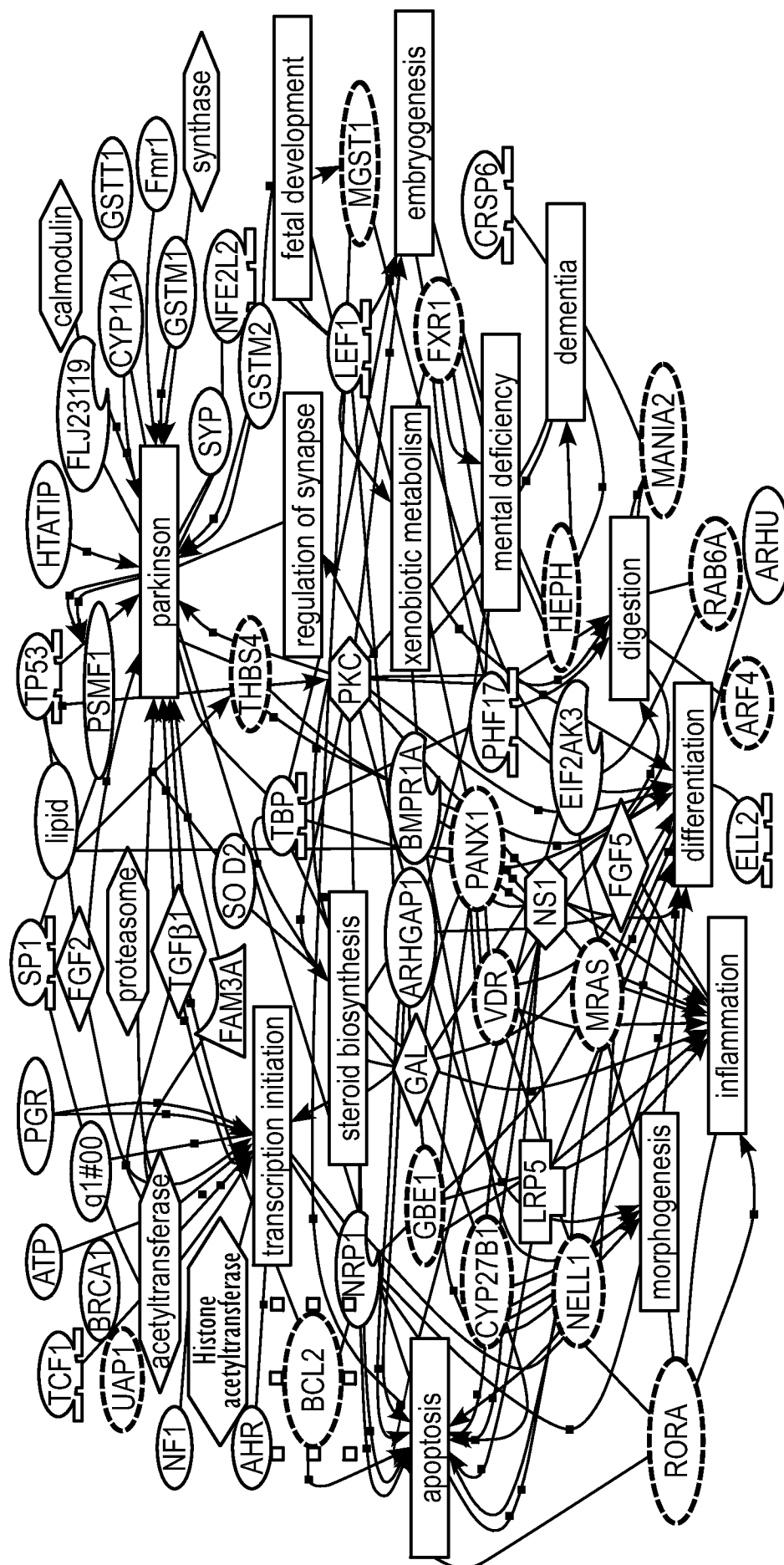


FIG. 2A

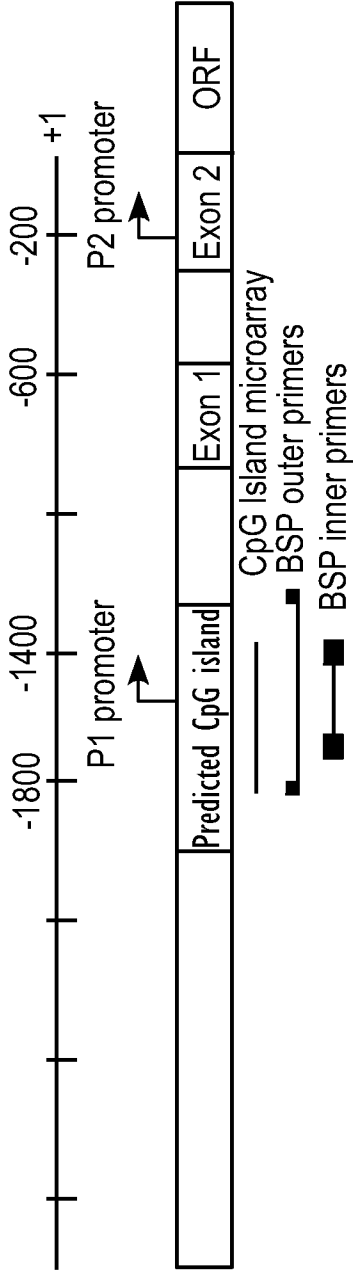


FIG. 2B

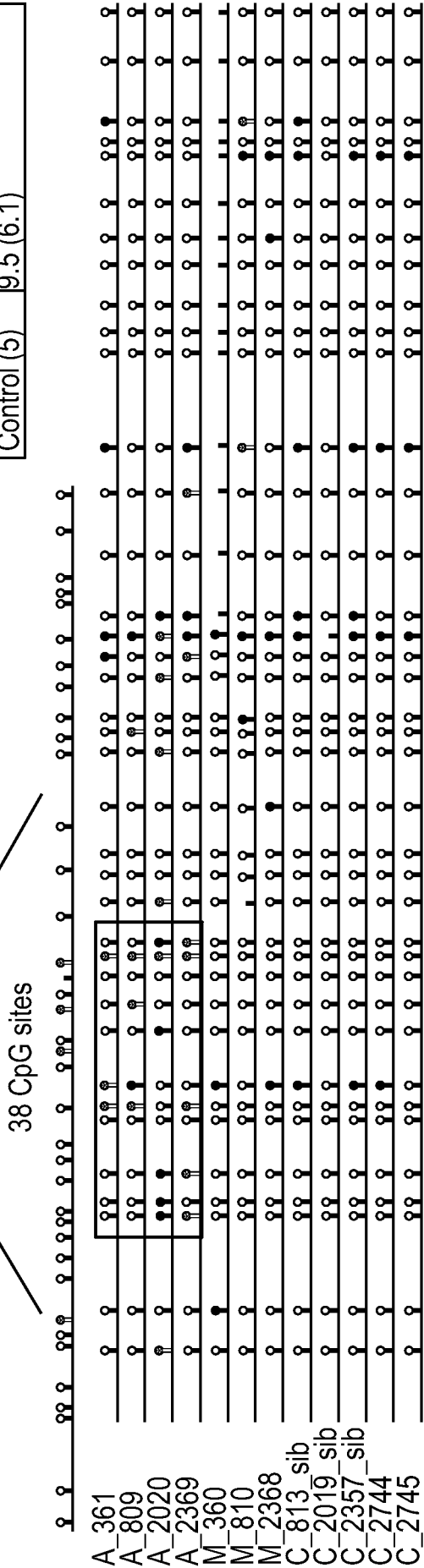


FIG. 3A

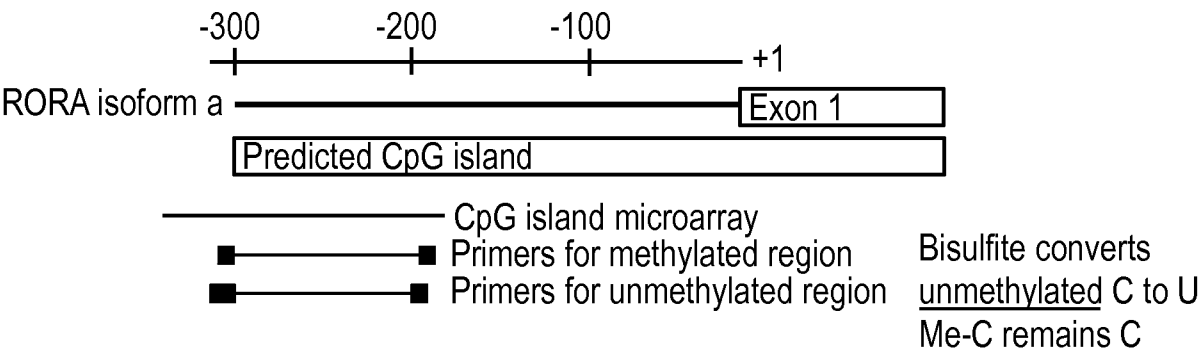
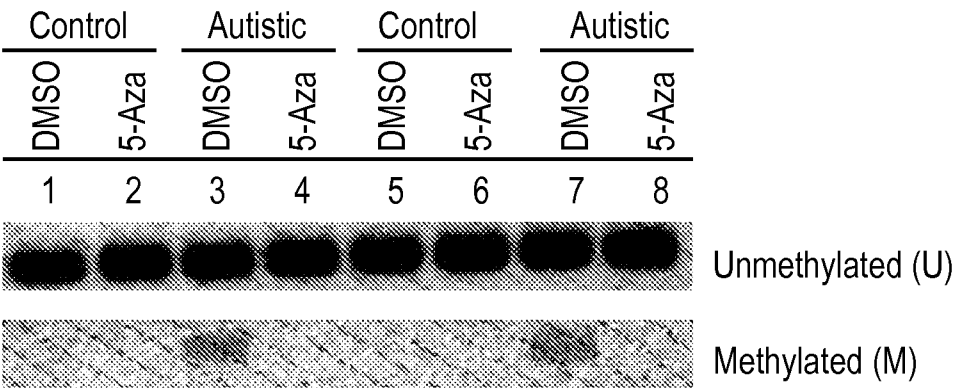
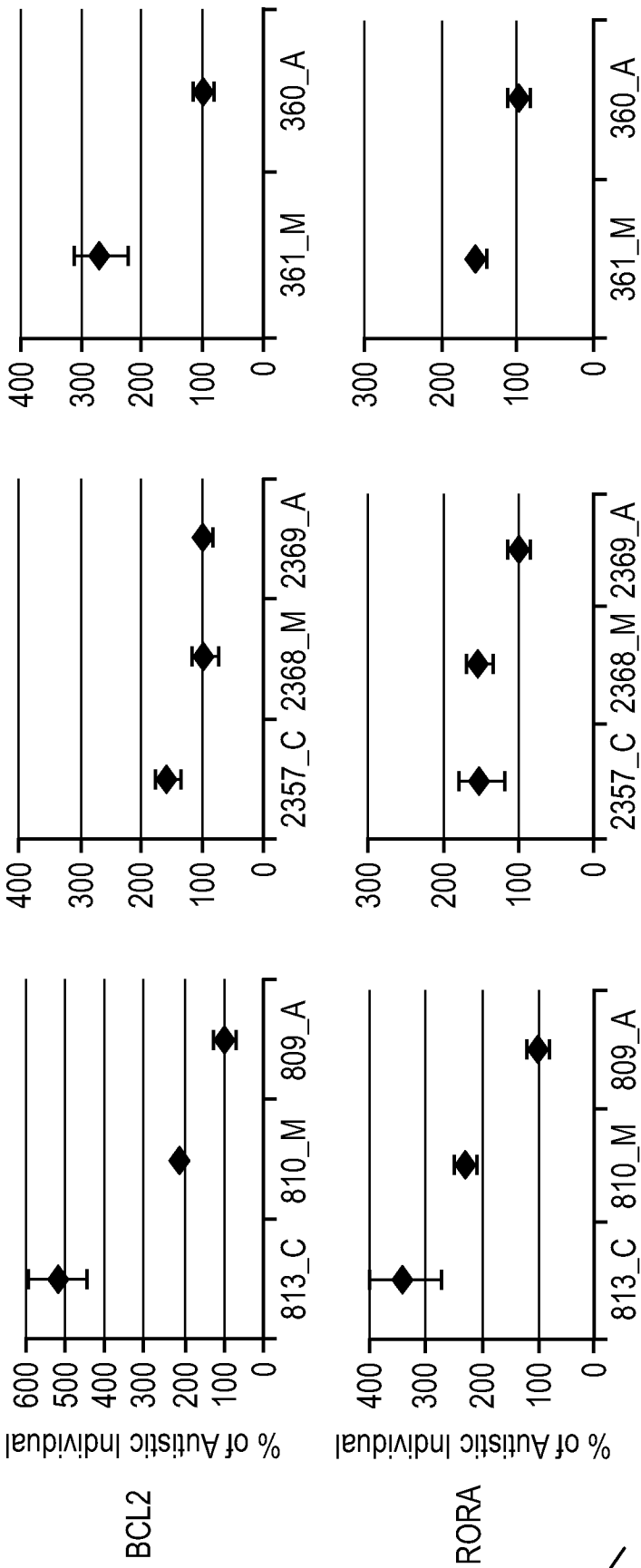


FIG. 3B



5/8

FIG. 4A



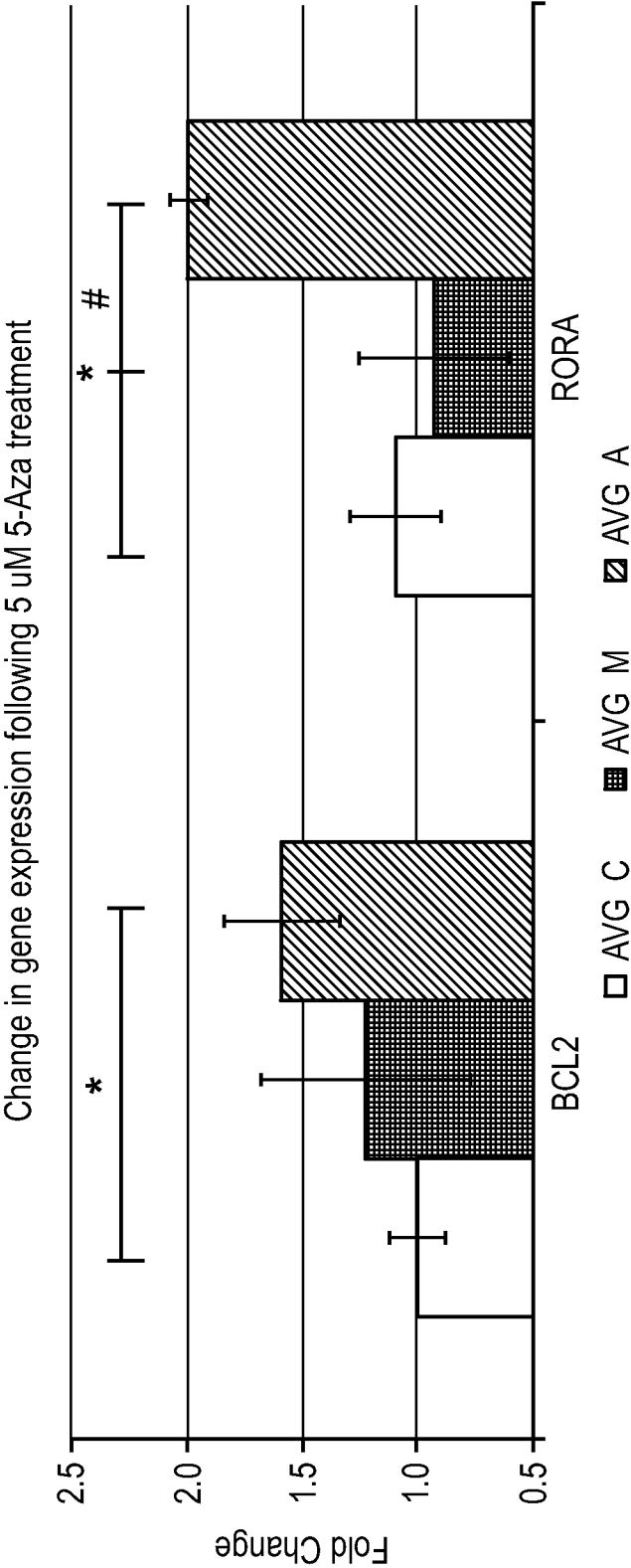


FIG. 4B

FIG. 5

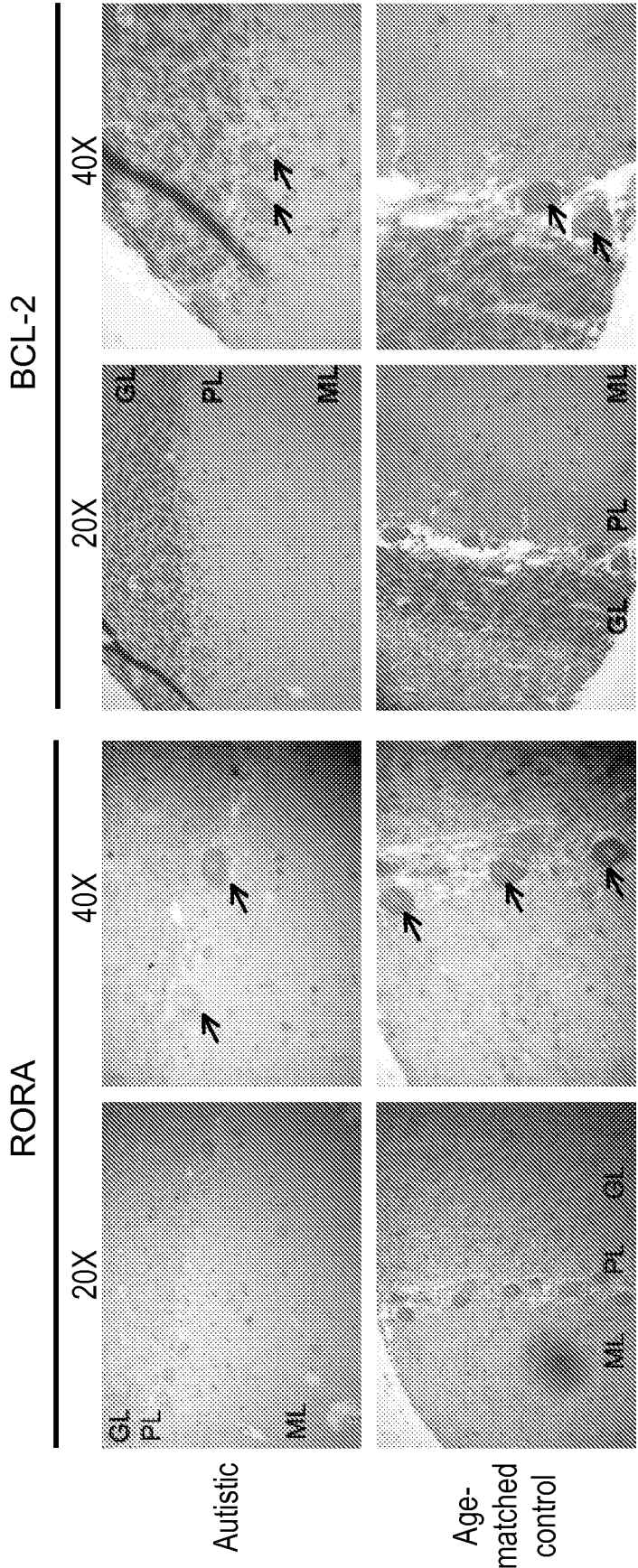


FIG. 6A

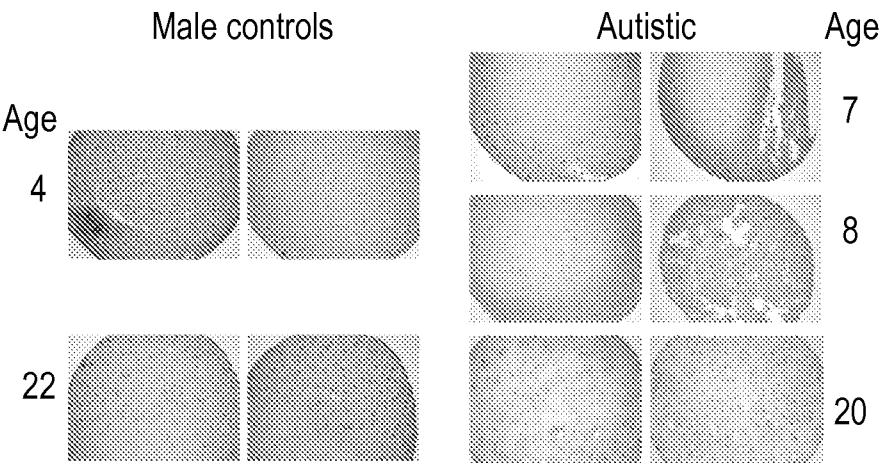
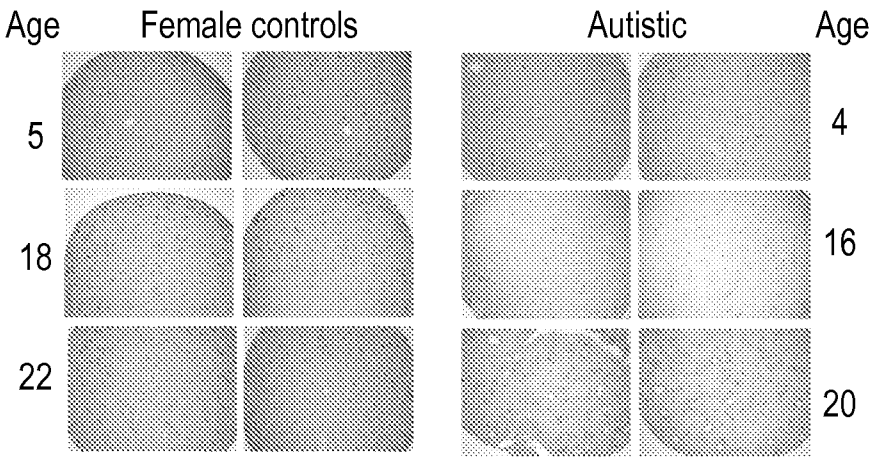


FIG. 6B



INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 11/31450

A. CLASSIFICATION OF SUBJECT MATTER

IPC(8) - C40B 30/04, C12Q 1/68 (2011.01)

USPC - 506/9, 435/6

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

USPC - 506/9, 435/6, 506/17, 506/7

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

PubWEST - PGPB,USPT,USOC,EPAB,JPAB; Dialog Classic Files - 654, 652, 349, 348, 35, 65, 155; USPTO Web Page; Search terms - oligonucleotide array, marker binding specificity, BCL2, RORA, ZDHHC9, DMD, UTP14A, BCORL1, CpG island methylation

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X - Y	US 2010/0029009 A1 (STEVENSON et al.) 04 February 2010 (04.02.2010) para [0019], [0029], [0034], [0062], [0069], [0073], [0074], [0080], [0090], [0096], [0107], abstract, Tables 1 and 2.	1, 3/1 ----- 2, 3/2
X	US 2009/0253583 A1 (YOGANATHAN) 08 October 2009 (08.10.2009) para [0019], [0077], [0100], [0172]-[0181]	1
Y	US 2009/0053706 A1 (LAIRD et al.) 26 February 2009 (26.02.2009) para [0003], [0014], [0015], [0028], [0070], [0120], [0121], [0267]	2, 3/2

☐ Further documents are listed in the continuation of Box C. ☐

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

17 August 2011 (17.08.2011)

Date of mailing of the international search report

26 AUG 2011

Name and mailing address of the ISA/US

Mail Stop PCT, Attn: ISA/US, Commissioner for Patents
P.O. Box 1450, Alexandria, Virginia 22313-1450

Facsimile No. 571-273-3201

Authorized officer:

Lee W. Young

PCT Helpdesk: 571-272-4300
PCT OSP: 571-272-7774

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 11/31450

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

Group I: Claims 1-3, drawn to an array comprising a plurality of different oligonucleotides with specificity to at least one of the genes contained in Table 1, Table 2, Table 3, or a combination thereof.

Group II: Claims 4-10 and 31-34, drawn to method for screening a subject for a neurological disease or disorder.

Group III: Claims 11-17 and 31-34, drawn to methods for determining differential DNA methylation profile for a autism spectrum disorder or for distinguishing different phenotypes of an autism spectrum disorder.

—please see continuation on extra sheet—

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
1-3

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
- ☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
- ☐ No protest accompanied the payment of additional search fees.

Continuation of Box No. III Observations where unity of invention is lacking

Group IV: Claims 18-21 and 31-34, drawn to a method of assessing the efficacy of a treatment in a subject having an autism spectrum disorder.

Group V: Claims 22-25 and 31-34, drawn to a method of identifying a DNA methylation profile responsive to administration of a therapeutic treatment.

Group VI: Claims 26-30 and 31-34, drawn to a method for conducting drug discovery.

Group VII: Claims 35-36, drawn to a kit and computer-readable medium.

The inventions listed as Groups I-VII do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons:

The special technical feature of the inventions listed as Groups I, II, IV, V, and VII is a gene set forth in Table 1, Table 2, Table 3. This special technical feature fails to provide a contribution over the prior art, as evidenced by US 2010/0029009 A1 to Stevensen et al. (published 4 February 2010). Stevensen discloses genes set forth in Applicant's Table 2 (para [0032] - "DMD"; Table 1 - "ZDHHC9"; Table 2 - "UTP14A" and "BCORL1") and teaches that the genes may be used in determining predisposition toward or diagnosis of an autism spectrum disorder (ASD) (para [0082]). Stevensen further teaches a method for determining predisposition to or diagnosis of autism spectrum disorder in an individual by determining the methylation level of a plurality of the genes (para [0027]-[0030]). In the absence of a contribution over the prior art, the shared technical feature is not a shared special technical feature.

The special technical feature of the inventions listed as Groups I, III, and V is an array or microarray for determining the methylation level of genes associated with autism. This special technical feature fails to provide a contribution over the prior art, as evidenced by Stevensen.

Stevensen teaches a microarray (para [0034], para [0072] - "methylation microarray"; para [0113] - "DNA Methylation array") that can comprise probes specific for genes associated with an autism spectrum disorder (para [0070], [0081]). In the absence of a contribution over the prior art, the shared technical feature is not a shared special technical feature.

The special technical feature of the inventions listed as Groups II, III, IV, V, VI, and VII is the assessment of DNA methylation profiles of genes associated with a neurological disorder, where the neurological disorder is an autism spectrum disorder. This special technical feature fails to provide a contribution over the prior art, as evidenced by Stevensen. Stevensen discloses a method of determining DNA methylation levels of genes associated with autism (para [0070], [0072]), wherein the method comprises determining and comparing the degree of methylation of an experimental sample to that of a known control (para [0073]). In the absence of a contribution over the prior art, the shared technical feature is not a shared special technical feature.

Further, the special technical feature of the invention listed as Group IV is the method step of assessing the efficacy of a treatment. This special technical feature is not shared by the inventions of Groups I-III and V-VII. The special technical feature of the invention listed as Group V is the method step of identifying a differential DNA methylation profile responsive to the administration of a therapeutic treatment. This special technical feature is not shared by the inventions of Groups I-IV and VI-VII. The special technical feature of the invention listed as Group VI is the method step of comparing the DNA methylation profile of a subject that was subjected to behavioral therapies to the DNA methylation profile of a subject that has been given a small molecule test agent. This special technical feature is not shared by the inventions of Groups I-V and VII. The special technical feature of the invention listed as Group VII is a computer program or database. This special technical feature is not shared by the inventions of Groups I-VI.

Unity of invention exists only when the same or corresponding technical feature is shared by the claimed inventions. With out a shared special technical feature, the inventions of Groups I -VII lack unity with one another.