(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
6 October 2016 (06.10.2016)

WIPO | PCT

(10) International Publication Number
WO 2016/160600 A1

(51) International Patent Classification:
C12Q 1/68 (2006.01) *G01N 33/50* (2006.01)
C40B 30/04 (2006.01) *G01N 33/53* (2006.01)
G01N 33/48 (2006.01) *G01N 33/68* (2006.01)

(21) International Application Number: PCT/US2016/024314

(22) International Filing Date: 25 March 2016 (25.03.2016)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/139,580 27 March 2015 (27.03.2015) US

(71) Applicant: THE JOHNS HOPKINS UNIVERSITY [US/US]; 3400 N. Charles Street, Baltimore, MD 21218 (US).

(72) Inventors: FEINBERG, Andrew, P.; 3400 N. Charles Street, Baltimore, MD 21218 (US). FALLIN, Daniele, M.; 3400 N. Charles Street, Baltimore, MD 21218 (US).

(74) Agents: HAILE, Lisa, A. et al.; DLA Piper LLP (US), 4365 Executive Drive, Suite 1100, San Diego, CA 92121-2133 (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, BN, 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, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, 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, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, 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, KM, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*

— *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))*

Published:

— *with international search report (Art. 21(3))*

[Continued on next page]

(54) **Title**: METHOD OF IDENTIFYING RISK FOR AUTISM

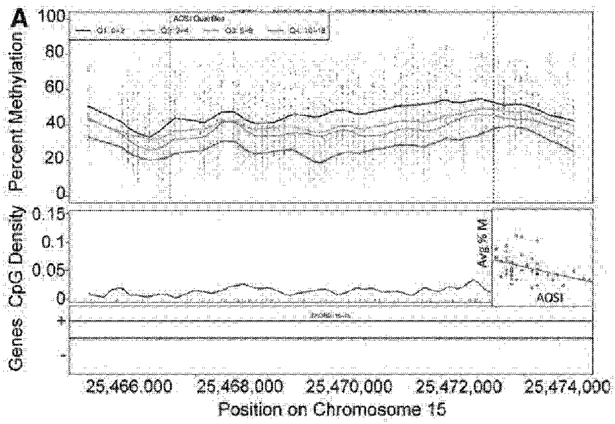

FIG. 1A

(57) **Abstract**: Disclosed herein is a method for determining risk of autism spectrum disorder (ASD) in an offspring subject. The method includes analyzing DNA methylation status in a sample containing sperm from the prospective paternal parent, wherein a methylation pattern that is different from the pattern found in a sample not associated with ASD, is indicative of a risk of ASD in the offspring.

WO 2016/160600 A1

— *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))*

METHOD OF IDENTIFYING RISK FOR AUTISM

RELATED APPLICATION DATA

[0001[This application claims the benefit of priority under 35 U.S.C. §119(e) of U.S. Provisional Patent Application Serial No. 62/139,580, filed March 27, 2015, the entire contents of which is incorporated herein by reference in its entirety.

STATEMENT OF GOVERNMENT SUPPORT

[0002[This invention was made in part with government support under Grant No. R01 ES017646 and R01 ES16443, both awarded by the National Institutes of Health. The United States government has certain rights in this invention.

BACKGROUND OF THE INVENTION

FIELD OF THE INVENTION

[0003[The invention relates generally to methylation based testing and more specifically to sperm DNA methylation as a predictive marker of autism risk.

BACKGROUND INFORMATION

[0004[Epigenetics is the study of non-sequence information of chromosome DNA during cell division and differentiation. The molecular basis of epigenetics is complex and involves modifications of the activation or inactivation of certain genes. Additionally, the chromatin proteins associated with DNA may be activated or silenced. Epigenetic changes are preserved when cells divide. Most epigenetic changes only occur within the course of one individual organism's lifetime, but some epigenetic changes are inherited from one generation to the next.

[0005[One example of an epigenetic mechanism is DNA methylation (DNAm), a covalent modification of the nucleotide cytosine. In particular, it involves the addition of methyl groups to cytosine nucleotides in the DNA, to convert cytosine to 5-methylcytosine. DNA methylation plays an important role in determining whether some genes are expressed or not. Abnormal DNA methylation is one of the mechanisms known to underlie the changes observed with aging and development of a host of disorders including many cancers, metabolic disorders and the like.

[0006[Epigenetic mechanisms such as altered DNA methylation, have been suggested to play a role in autism, beginning with the classical association of Prader-Willi syndrome, an imprinting disorder, with autistic features. Autism spectrum disorder (ASD) is a complex neurodevelopmental disorder characterized by deficits in communication and social behaviors, as well as stereotypic movements. Recent estimates from the Autism and Developmental Disabilities Monitoring (ADDM) Network report the prevalence of ASD among eight year old US children to be 1 in 68. The recurrence rate of ASD among children with an older affect sibling is estimated at 18.7% (95% CI = 13.3% to 25.5%) in the Baby Siblings Research Consortium, representing 10-fold enrichment over the general population. Neuroanatomy evidence, teratogen exposure studies, and very early life behavior differences suggest that ASD neurobiology likely begins during the *in utero* period. Intervention research demonstrates early identification and diagnosis lead to better treatment outcomes for children with ASD, thus identification of prenatal or early life ASD risk factors are critical to disease prevention or mitigation.

[0007[The etiology of ASD is incompletely characterized, and recent heritability estimates suggest approximately equal contribution from genetic and environmental risk factors. Several large genome-wide studies have identified *de novo* and inherited genetic risk factors for ASD, yet there appears to be high genetic heterogeneity and much of the overall ASD risk has not yet been explained. Similarly, evidence of prenatal toxicant risk factors for ASD is growing, but inconsistent to date. Parental characteristics may be associated with ASD risk; numerous studies have observed associations between advanced paternal age, maternal age, or both. Parental age, particularly paternal age, has been implicated in d*e novo* genetic findings for ASD and has been associated with brain expression of particular genes in offspring. Other paternal risk factors might also be important, however. *In vitro* fertilization by intracytoplasmic sperm injection is associated with an increased frequency of autism that is specific for method, with a 4.6 fold increased risk after surgical sperm extraction and >9-fold if also associated with preterm birth. Consideration of combined genetic, environmental and demographic factors is likely to yield further insights into ASD etiology.

[0008[Epigenetics, defined as genomic modifications heritable during cell division other than DNA sequence *per se*, such as DNA methylation (DNAm), represents an important intersection of genetic and environmental influences that may be particularly relevant for ASD. DNAm alterations have been observed in multiple regions of post mortem brains of individuals with ASD compared to controls and between blood samples from discordant ASD twins. Epigenetic changes to the developing fetus may underscore risk of ASD, and nutritional deprivation can impact the male germline with epigenetic changes to the offspring.

SUMMARY OF THE INVENTION

[0009[In one embodiment, the invention provides a method for determining risk of autism spectrum disorder (ASD) in an offspring subject comprising analyzing DNA methylation status in a sample containing sperm from the prospective paternal parent, wherein a methylation pattern that is different from the pattern found in a sample not associated with ASD, is indicative of a risk of ASD in the offspring. In a preferred aspect, the sample is a semen sample. In a preferred aspect, the subject is a human.

[00010[In another embodiment, the invention provides a method for determining whether a subject has or is at risk of having autism spectrum disorder (ASD) comprising analyzing DNA methylation status in a DNA sample of the subject, wherein a methylation pattern that is different from the pattern found in a sample not associated with ASD, is indicative of a risk of ASD in the subject. In one aspect, the sample is a blood or tissue sample. In a preferred aspect, the subject is a human. In the method of the invention, risk is assessed as a score relative to low, moderate or high risk.

[00011[In the methods of the invention the step of determining a methylation status comprises determining the methylation status of differentially methylated regions (DMRs) in the DNA, or a gene or regulatory region thereof that is associated with ASD. While not wishing to be limited, the methylation status is performed by one or more techniques selected from the group consisting of a nucleic acid amplification, polymerase chain reaction (PCR), methylation specific PCR, bisulfite sequencing (e.g., capture or whole genome), pyrosequencing, single-strand conformation polymorphism (SSCP) analysis, restriction analysis, microarray technology, including bead microarray technology, and proteomics. Analysis of methylation status can be done by performing bead array analysis or comprehensive high-through array-based relative methylation (CHARM) analysis on a sample of labeled, digested genomic DNA, for example. The methylation status may show as hypo- or hyper-methylation.

[00012[In one aspect, the DMRs are associated with genes for neurogenesis and/or neuronal development. For illustrative purposes, the DMRs are selected from the DMRs in Table 2, Table, 6, Table 7, and Table 8, herein. In one aspect, the DMR is associated with Prader-Willi syndrome. For example, the DMR may reside on chromosome 15.

[00013[In one aspect, methylation status is determined in a gene or regulatory region thereof that is associated with ASD, for example, those set forth in Table 2, Table, 6, Table 7, and Table 8, herein. In particular embodiments, the gene or regulatory region thereof includes one or more of SNORD115-15, SNORD115-11, SNORD115-17, SMYD3, MUC17, RBM19, FAM13A, GRP125, WDR1 or a combination thereof.

[00014[In another aspect, the invention provides a plurality of nucleic acid sequences that selectively hybridize to a nucleic acid sequence, such as a differentially methylated region (DMR) set forth in Table 2, Table, 6, Table 7, and Table 8, or complement thereof. In embodiments, each of the plurality of nucleic acid sequences is about 10-55 base pairs in length.

[00015[In still another aspect, the invention provides a microarray which may be used to analyze methylation status in a sample from subject having or at risk of having ASD. In one embodiments the microarray includes one or more of the plurality of nucleic acid sequences which selectively hybridize to a differentially methylated region (DMR) set forth in Table 2, Table, 6, Table 7, and Table 8, or complement thereof.

[00016[In yet another aspect, the invention provides a method of determining methylation status of genomic DNA isolated from a cell. The invention provides performing an assay utilizing the microarray of the invention. In one embodiment, the assay is comprehensive high-throughput array for relative methylation (CHARM) analysis on a sample of labeled, digested genomic DNA isolated from the cell. In another aspect, the assay utilizes a bead array format.

[00017[In another aspect, the invention provides a kit for determining whether a subject has, or is at risk of having or inheriting, ASD. The kit includes a reagent for determining methylation status of a nucleic acid sequence, wherein the nucleic acid sequence is a differentially methylated region (DMR) associated with ASD, or a gene or regulatory region thereof which is associated with ASD; and instructions for use of the reagent. In embodiments, the DMR is one or more DMRs set forth in Table 2, Table, 6, Table 7, and Table 8. In embodiments the gene or regulatory region thereof is one or more as set forth in Table 2, Table, 6, Table 7, and Table 8. In embodiments, the reagent is an oligonucleotide probe, primer, or primer pair, or combination thereof, capable of selectively hybridizing to a gene or regulatory region thereof, or DMR associated with ASD, with or without prior bisulfite treatment of the DMR. Optionally the kit includes one or more detectable labels. Further, the kit may include a plurality of oligonucleotide probes, primers, or primer pairs, or combinations thereof, capable of binding to DMRs with or without prior bisulfite treatment of the DMRs. The kit may also include an oligonucleotide primer pair that hybridizes under stringent conditions to all or a portion of the DMRs only after bisulfite treatment.

BRIEF DESCRIPTION OF THE DRAWINGS

[00018[Figures 1A-1D are graphical representations of methylation plots for the top 4 statistical DMRs identified using CHARM and 12-month score in embodiments of the present invention.

[00019[Figure 2 is a graphical representation depicting validation of CHARM AOSI DMR values using 450k data from overlapping samples (n=30).

[00020[Figure 3 is a graphical representation showing the relationship between linear regression beta coefficients of the original bump hunting model (Beta1) and the model including principal components of race from ancestry analysis of GWAS SNP data (Beta2).

[00021[Figure 4 sets forth Table 1 relating to bivariate associations of AOSI score with demographic and laboratory variables for the EARLI study population.

[00022[Figure 5 sets forth Table 2 relating to candidate differentially methylated regions (DMRs) associated with 12-month total AOSI score, an indicator of autism risk.

[00023[Figure 6 sets forth Table 3 relating to gene ontology, biological processes enriched in differentially methylated regions associated with 12-month AOSI score.

[00024[Figure 7 sets forth Table 4 relating to association between SVs and variables for CHARM samples.

[00025[Figure 8 sets forth Table 5 relating to association between SVs and variables for 450k samples.

[00026[Figure 9 sets forth Table 6 relating to the top 193 DMRs associated with autism as determined by an embodiment of the invention.

[00027[Figure 10 sets forth Table 7 relating to validation of 450k probes.

[00028[Figure 11 sets forth Table 8 relating to overlap of AOSI DMRs.

DETAILED DESCRIPTION OF THE INVENTION

[00029[The present invention is based on the seminal discovery of the relationship of paternal sperm DNA methylation with autism risk in offspring, by examining an enriched-risk cohort of fathers of autistic children.

[00030[As described in the Examples herein, genome-wide DNAm was examined in paternal semen biosamples obtained from an ASD enriched-risk pregnancy cohort, the Early Autism Risk Longitudinal Investigation (EARLI) cohort, to estimate associations between sperm DNAm and prospective ASD development, using a 12-month ASD symptoms assessment, the Autism Observation Scale for Infants (AOSI). Methylation data from 44 sperm samples were analyzed and run on the CHARM 3.0™ array, which contains over 4 million probes (over 7 million CpG sites), including 30 samples also run on the Illumina Infinium HumanMethylation450 ™ (450k) BeadChip ™ platform (~485,000 CpG sites). Associated regions were also examined in an independent sample of postmortem human brain ASD and control samples for which Illumina 450K™ DNA methylation data were available.

[00031[Before the present compositions and methods are described, it is to be understood that this invention is not limited to particular compositions, methods, and experimental conditions described, as such compositions, methods, and conditions may vary. It is also to be understood that the terminology used herein is for purposes of describing particular embodiments only, and is not intended to be limiting, since the scope of the present invention will be limited only in the appended claims.

[00032[As used in this specification and the appended claims, the singular forms "a", "an", and "the" include plural references unless the context clearly dictates otherwise. Thus, for example, references to "the method" includes one or more methods, and/or steps of the type described herein which will become apparent to those persons skilled in the art upon reading this disclosure and so forth.

[00033[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. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the invention, the preferred methods and materials are now described.

[00034[Using region-based statistical approaches, 193 differentially methylated regions (DMRs) were identified in paternal sperm with a family-wise empirical p value (FWER) < 0.05 associated with performance on the Autism Observational Scale for Infants (AOSI) at 12 months in offspring. The DMRs clustered near genes involved in developmental processes as set forth in Table 2, Table, 6, Table 7, and Table 8, including many genes in the *SNORD* family, within the Prader-Willi syndrome gene cluster. These results were consistent among the 75 probes on the Illumina 450K™ array that cover AOSI-associated DMRs from CHARM. Further, 18 of 75 (24%) 450K array probes showed consistent differences in the cerebellums of autistic individuals compared to controls.

[00035[These data suggest that epigenetic differences in paternal sperm may contribute to autism risk in offspring, and provide evidence that directionally consistent, potentially related epigenetic mechanisms may be operating in the cerebellum of individuals with autism.

[00036[In a high-risk autism spectrum disorder (ASD) birth cohort, paternal sperm genome-wide DNA methylation patterns are associated with 12-month ASD-related phenotypes. Observations were validated in a partially independent set of fathers using a second method of DNA methylation assessment. Regions of altered sperm DNA methylation in fathers associated with infant ASD phenotypes included overlapping genes previously associated with Prader-Willi syndrome. A significant subset of sperm DMRs associated with ASD phenotypes showed directionally consistent methylation changes in postmortem cerebellar tissue of ASD patients compared to controls.

[00037[Accordingly, in one aspect, the invention provides a method for determining risk of autism spectrum disorder (ASD) in an offspring subject. The method includes a) analyzing DNA methylation status in a sample containing sperm from the prospective paternal parent; and b) determining a deviation in the methylation status of (a) as compared to a corresponding normal sample not associated with ASD, wherein a methylation pattern that deviates from the pattern found in the corresponding sample not associated with ASD, is indicative of a risk of ASD in the offspring.

[00038[In another aspect, the invention provides a method for determining whether a subject has or is at risk of having autism spectrum disorder (ASD). The method includes a) analyzing DNA methylation status in a DNA sample of the subject; and b) determining a deviation in the methylation status of (a) as compared to a corresponding normal sample not associated with ASD, wherein a methylation pattern that is different from the pattern found in the corresponding sample not associated with ASD, is indicative of a subject having or being at risk of having ASD.

[00039[Autism, one ASD, is mostly diagnosed clinically using behavioral criteria because few specific biological markers are known for diagnosing the disease. Autism is a neuropsychiatric developmental disorder characterized by impaired verbal communication, non-verbal communication, and reciprocal social interaction. It is also characterized by restricted and stereotyped patterns of interests and activities, as well as the presence of developmental abnormalities by 3 years of age. Autism-associated disorders, diseases or pathologies can comprise any metabolic, immune or systemic disorders; gastrointestinal disorders; epilepsy; congenital malformations or genetic syndromes; anxiety, depression, or AD/HD; or speech delay and motor in-coordination.

[00040[Autism spectrum disorder (ASD) is defined by impairments in verbal and non-verbal communication, social interactions, and repetitive and stereotyped behaviors. In addition to these core deficits, previous reports indicate that the prevalence of gastrointestinal symptoms ranges widely in individuals with ASD, from 9 to 91%. Macroscopic and histological observations in ASD include findings of ileo-colonic lymphoid nodular hyperplasia (LNH), enterocolitis, gastritis and esophagitis. Associated changes in intestinal inflammatory parameters include higher densities of lymphocyte populations, aberrant cytokine profiles, and deposition of immunoglobulin (IgG) and complement C1q on the basolateral enterocyte membrane. Functional disturbances include increased intestinal permeability, compromised sulphoconjugation of phenolic compounds, deficient enzymatic activity of disaccharidases, increased secretin-induced pancreatico-biliary secretion, and abnormal Clostridia taxa. Some children placed on exclusion diets or treated with the antibiotic vancomycin are reported to improve in cognitive and social function.

[00041[In various embodiments, a genome is present in a biological sample taken from a subject. The biological sample can be virtually any biological sample, particularly a sample that contains DNA from the subject. The biological sample can be a semen or tissue sample which contains about 1 to about 10,000,000, about 1000 to about 10,000,000, or about 1,000,000 to about 10,000,000 cells. However, it is possible to obtain samples that contain smaller numbers of cells, even a single cell in embodiments that utilize an amplification protocol such as PCR. The sample need not contain any intact cells, so long as it contains sufficient biological material (e.g., protein or genetic material, such as RNA or DNA) to assess methylation status of the one or more DMRs.

[00042[In some embodiments, a biological or tissue sample can be drawn from any tissue that includes cells with DNA. A biological or tissue sample may be obtained by surgery, biopsy, swab, stool, or other collection method. In some embodiments, the sample is derived from blood, plasma, serum, lymph, nerve-cell containing tissue, cerebrospinal fluid, biopsy material, tumor tissue, bone marrow, nervous tissue, skin, hair, tears, fetal material, amniocentesis material, uterine tissue, saliva, feces, or sperm. Methods for isolating PBLs from whole blood are well known in the art.

[00043[As disclosed above, the biological sample can be a blood sample. The blood sample can be obtained using methods known in the art, such as finger prick or phlebotomy. Suitably, the blood sample is approximately 0.1 to 20 ml, or alternatively approximately 1 to 15 ml with the volume of blood being approximately 10 ml.

[00044[In the present invention, the subject is typically a human but also can be any mammal, including, but not limited to, a dog, cat, rabbit, cow, bird, rat, horse, pig, or monkey.

[00045[To examine DNAm on a genome-wide scale, comprehensive high-throughput array-based relative methylation (CHARM) analysis, which is a microarray-based method agnostic to preconceptions about DNAm, including location relative to genes and CpG content is carried out. The resulting quantitative measurements of DNAm, denoted with M, are log ratios of intensities from total (Cy3) and McrBC-fractionated DNA (Cy5): positive and negative M values are quantitatively associated with methylated and unmethylated sites, respectively.

[00046[While the present invention exemplifies the CHARM assay for detection of methylation, in fact numerous methods for analyzing methylation status (hypomethylation or hypermethylation) may be utilized. In various embodiments, the determining of methylation status in the methods of the invention is performed by one or more techniques selected from the group consisting of a nucleic acid amplification, polymerase chain reaction (PCR), methylation specific PCR, bisulfite pyrosequenceing, single-strand conformation polymorphism (SSCP) analysis, restriction analysis, microarray technology, and proteomics. As illustrated in the Examples herein, analysis of methylation can be performed by bisulfite genomic sequencing. Bisulfite treatment modifies DNA converting unmethylated, but not methylated, cytosines to uracil. Bisulfite treatment can be carried out using the METHYLEASY™ bisulfite modification kit (Human Genetic Signatures).

[00047[In some embodiments, bisulfite pyrosequencing, which is a sequencing-based analysis of DNA methylation that quantitatively measures multiple, consecutive CpG sites individually with high accuracy and reproducibility, may be used. Exemplary primers for such analysis are set forth in in the present disclosure.

[00048[It will be recognized that depending on the site bound by the primer and the direction of extension from a primer, that the primers listed above can be used in different pairs. Furthermore, it will be recognized that additional primers can be identified within the

DMRs, especially primers that allow analysis of the same methylation sites as those analyzed with primers that correspond to the primers disclosed herein.

[00049[Bisulfite treatment can be carried out using the CpG Genome DNA Modification™ kit (Intergen, Purchase, N.Y.) with the following modifications of the manufacturer's protocol: denatured genomic DNA (4 ug) can be incubated at 55 degrees C in the dark overnight in 1100 ul of freshly prepared Reagent I, with subsequent column purification with the QIAquick PCR™ purification kit (Qiagen). Purified DNA can be treated at 37 degrees C for 15 min with freshly prepared 3 M NaOH to a final concentration of 0.3 M NaOH. Then the DNA can be precipitated with ethanol and dissolved in 40 ul of 10 mM Tris (pH 8)-1 mM EDTA for nested PCR. PCR products were purified on 2% agarose gels for direct sequencing as described above. The annealing temperature was 55 degrees C. For sequencing individual clones, the PCR products can be subcloned into a TA Cloning vector (Invitrogen, Carlsbad, Calif.) according to the manufacturer's instructions, and a series of clones, such as 10-15 clones, can be selected for sequencing.

[00050[PCR products can be purified using the QIAEX II™ gel extraction kit (Qiagen) and directly sequenced with an ABI Prism 377 DNA™ sequencer using the BigDye™ Terminator Cycle Sequencing kit following the manufacturer's protocol (PE Applied Biosystems, Foster City, Calif.).

[00051[Altered methylation can be identified by identifying a detectable difference in methylation. For example, hypomethylation can be determined by identifying whether after bisulfite treatment a uracil or a cytosine is present a particular location. If uracil is present after bisulfite treatment, then the residue is unmethylated. Hypomethylation is present when there is a measurable decrease in methylation.

[00052[In an alternative embodiment, the method for analyzing methylation of the DMR can include amplification using a primer pair specific for methylated residues within a DMR. In these embodiments, selective hybridization or binding of at least one of the primers is dependent on the methylation state of the target DNA sequence (Herman et al., *Proc. Natl. Acad. Sci. USA*, 93:9821 (1996)). For example, the amplification reaction can be preceded by bisulfite treatment, and the primers can selectively hybridize to target sequences in a manner that is dependent on bisulfite treatment. For example, one primer can selectively bind to a target sequence only when one or more base of the target sequence is altered by bisulfite treatment, thereby being specific for a methylated target sequence.

[00053[Methods using an amplification reaction, for example methods above for detecting hypomethylation or hyprmethylation of one or more DMRs, can utilize a real-time detection amplification procedure. For example, the method can utilize molecular beacon technology (Tyagi et al., Nature Biotechnology, 14: 303 (1996)) or Taqman™ technology (Holland et al., Proc. Natl. Acad. Sci. USA, 88:7276 (1991)).

[00054[Also methyl light (Trinh et al., Methods 25(4):456-62 (2001), incorporated herein in its entirety by reference), Methyl Heavy (Epigenomics, Berlin, Germany), or SNuPE (single nucleotide primer extension) (see e.g., Watson et al., Genet Res. 75(3):269-74 (2000)) Can be used in the methods of the present invention related to identifying altered methylation of DMRs.

[00055[Other methods are known in the art for determining methylation status of a DMR, including, but not limited to, array-based methylation analysis and Southern blot analysis. Additionally, as mentioned above, methyl light, methyl heavy, and array-based methylation analysis can be performed, by using bisulfate treated DNA that is then PCR-amplified, against microarrays of oligonucleotide target sequences with the various forms corresponding to unmethylated and methylated DNA.

[00056[As used herein, the term "selective hybridization" or "selectively hybridize" refers to hybridization under moderately stringent or highly stringent physiological conditions, which can distinguish related nucleotide sequences from unrelated nucleotide sequences.

[00057[As known in the art, in nucleic acid hybridization reactions, the conditions used to achieve a particular level of stringency will vary, depending on the nature of the nucleic acids being hybridized. For example, the length, degree of complementarity, nucleotide sequence composition (for example, relative GC:AT content), and nucleic acid type, for example, whether the oligonucleotide or the target nucleic acid sequence is DNA or RNA, can be considered in selecting hybridization conditions. An additional consideration is whether one of the nucleic acids is immobilized, for example, on a filter. Methods for selecting appropriate stringency conditions can be determined empirically or estimated using various formulas, and are well known in the art (*see*, *e.g.*, Sambrook et al., supra, 1989).

[00058[An example of progressively higher stringency conditions is as follows: 2X SSC/0.1% SDS at about room temperature (hybridization conditions); 0.2X SSC/0.1% SDS at about room temperature (low stringency conditions); 0.2X SSC/0.1% SDS at about 42°C (moderate stringency conditions); and 0.1X SSC at about 68°C (high stringency conditions). Washing can be carried out using only one of these conditions, for example, high stringency conditions, or each of the conditions can be used, for example, for 10 to 15 minutes each, in the order listed above, repeating any or all of the steps listed.

[00059[The degree of methylation in the DNA associated with the DMRs being assessed, may be measured by fluorescent *in situ* hybridization (FISH) by means of probes which identify and differentiate between genomic DNAs, associated with the DMRs being assessed, which exhibit different degrees of DNA methylation. FISH is described, for example, in de Capoa et al. (*Cytometry*. 31:85-92, 1998) which is incorporated herein by reference. In this case, the biological sample will typically be any which contains sufficient whole cells or nuclei to perform short term culture. Usually, the sample will be a sample that contains 10 to 10,000, or, for example, 100 to 10,000, whole cells.

[00060[The term "nucleic acid molecule" is used broadly herein to mean a sequence of deoxyribonucleotides or ribonucleotides that are linked together by a phosphodiester bond. As such, the term "nucleic acid molecule" is meant to include DNA and RNA, which can be single stranded or double stranded, as well as DNA/RNA hybrids. Furthermore, the term "nucleic acid molecule" as used herein includes naturally occurring nucleic acid molecules, which can be isolated from a cell, as well as synthetic molecules, which can be prepared, for example, by methods of chemical synthesis or by enzymatic methods such as by the polymerase chain reaction (PCR), and, in various embodiments, can contain nucleotide analogs or a backbone bond other than a phosphodiester bond.

[00061[The terms "polynucleotide" and "oligonucleotide" also are used herein to refer to nucleic acid molecules. Although no specific distinction from each other or from "nucleic acid molecule" is intended by the use of these terms, the term "polynucleotide" is used generally in reference to a nucleic acid molecule that encodes a polypeptide, or a peptide portion thereof, whereas the term "oligonucleotide" is used generally in reference to a nucleotide sequence useful as a probe, a PCR primer, an antisense molecule, or the like. Of course, it will be recognized that an "oligonucleotide" also can encode a peptide. As such, the different terms are used primarily for convenience of discussion.

[00062[A polynucleotide or oligonucleotide comprising naturally occurring nucleotides and phosphodiester bonds can be chemically synthesized or can be produced using recombinant DNA methods, using an appropriate polynucleotide as a template. In comparison, a polynucleotide comprising nucleotide analogs or covalent bonds other than phosphodiester bonds generally will be chemically synthesized, although an enzyme such as T7 polymerase can incorporate certain types of nucleotide analogs into a polynucleotide and, therefore, can be used to produce such a polynucleotide recombinantly from an appropriate template.

[00063[In various aspects of the invention, methylation status is converted to an M value. As used herein an M value, can be a log ratio of intensities from total (Cy3) and McrBC-fractionated DNA (Cy5): positive and negative M values are quantitatively associated with methylated and unmethylated sites, respectively.

[00064[In various aspects of the invention DMR may be hypermethylated or hypomethylated. Hypomethylation of a DMR is present when there is a measurable decrease in methylation of the DMR. In some embodiments, a DMR can be determined to be hypomethylated when less than 50% of the methylation sites analyzed are not methylated. Hypermethylation of a DMR is present when there is a measurable increase in methylation of the DMR. In some embodiments, a DMR can be determined to be hypermethylated when more than 50% of the methylation sites analyzed are methylated. Methods for determining methylation states are provided herein and are known in the art. In some embodiments methylation status is converted to an M value. As used herein an M value, can be a log ratio of intensities from total (Cy3) and McrBC-fractionated DNA (Cy5): positive and negative M values are quantitatively associated with methylated and unmethylated sites, respectively. M values are calculated as described in the Examples. In some embodiments, M values which range from −0.5 to 0.5 represent unmethylated sites as defined by the control probes, and values from 0.5 to 1.5 represent baseline levels of methylation.

[00065[In one embodiment of the invention, methylation density is determined for a region of nucleic acid, such as a DMR. Density may be used as an indication of production of an iPS cell, for example. A density of about 0.2 to 0.7, about 0.3 to 0.7 , 0.3 to 0.6 or 0.3 to 0.4, or 0.3, may be indicative of ASD (the calculated DNA methylation density is the number of methylated CpGs divided by the total number of CpGs sequenced for each sample). Methods for determining methylation density are well known in the art. For example, a method for determining methylation density of target CpG islands has been established by Luo et al. (*Analytical Biochemistry*, Vol. 387:2 2009, pp. 143-149). In the method, DNA microarray was prepared by spotting a set of PCR products amplified from bisulfite-converted sample DNAs. This method not only allows the quantitative analysis of regional methylation density of a set of given genes but also could provide information of methylation density for a large amount of clinical samples as well as use in the methods of the invention regarding iPS cell generation and detection. Other methods are well known in the art (*e.g.*, Holemon et al., *BioTechniques,* 43:5, 2007, pp. 683–693).

[00066[In another aspect, the present invention includes kits that are useful for carrying out the methods of the present invention. The components contained in the kit depend on a number of factors, including: the particular analytical technique used to detect methylation or measure the degree of methylation or a change in methylation, and the one or more DMRs or genes being assayed for methylation status.

[00067[Accordingly, the present invention provides a kit for determining a methylation status of one or more DMRs of the invention. In some embodiments, the one or more DMRs are selected from one or more of the sequences as set forth in Table 2, Table, 6, Table 7, and Table 8, or DMRs associated with Prader-Willi Syndrome, *e.g.,* on chromosome 15. The kit includes an oligonucleotide probe, primer, or primer pair, or combination thereof for carrying out a method for detecting methylation status, as discussed above. For example, the probe, primer, or primer pair, can be capable of selectively hybridizing to the DMR either with or without prior bisulfite treatment of the DMR. The kit can further include one or more detectable labels.

[00068[The kit can also include a plurality of oligonucleotide probes, primers, or primer pairs, or combinations thereof, capable of selectively hybridizing to the DMR with or without prior bisulfite treatment of the DMR. The kit can include an oligonucleotide primer pair that hybridizes under stringent conditions to all or a portion of the DMR only after bisulfite treatment. In one aspect, the kit can provide reagents for bisulfite sequencing or pyrosequencing. The kit can include instructions on using kit components to identify, for example, the presence of autism or increased risk for developing autism.

[00069[In a related embodiment, the kit may further include computer executable code and instructions for performing statistical analysis.

[00070[In another aspect, the invention provides a plurality of nucleic acid sequences that selectively hybridize to a nucleic acid sequence, such as a differentially methylated region (DMR) set forth in Table 2, Table, 6, Table 7, and Table 8, or complements thereof.

[00071[In another aspect, the invention provides a plurality of nucleic acid sequences that selectively hybridize to a nucleic acid sequence including a gene or promoter region thereof as set forth in Table 2, Table, 6, Table 7, and Table 8, or complements thereof. In some embodiments, the plurality of nucleic acid sequences selectively hybridize to one or more of the following genes or regulatory regions thereof: SNORD115-15, SNORD115-11, SNORD115-17, SMYD3, MUC17, RBM19, FAM13A, GRP125, WDR1 or a combination thereof, or complements thereof.

[00072[In still another aspect, the invention provides a microarray which may be used to analyze methylation status in a sample from subject having or at risk of having ASD. In one embodiments the microarray includes one or more of the plurality of nucleic acid sequences which selectively hybridize to a differentially methylated region (DMR) set forth in Table 2, Table, 6, Table 7, and Table 8, or complements thereof, or to one or more of the following genes or regulatory regions thereof: SNORD115-15, SNORD115-11, SNORD115-17, SMYD3, MUC17, RBM19, FAM13A, GRP125, WDR1 or a combination thereof, or complements thereof. One skilled in the art would appreciate the many techniques that are well known for attaching nucleic acids on a substrate that may be utilized along with the various types of substrates and configurations.

[00073[In a related embodiment, the microarray of the present invention may be included in the kit of the present invention optionally along with reagents for performing an array based assay.

[00074[Polynucleotides or nucleic acid sequences of the present invention, such as oligonucleotides, primers, probes, and the like may be of any suitable length. For example, one of skill in the art would understand what lengths are suitable for nucleic acid sequences to be used in an array or kit of the invention. Such molecules are typically from about 5 to 100, 5 to 50, 5 to 45, 5 to 40, 5 to 35, 5 to 30, 5 to 25, 5 to 20, or 10 to 20 nucleotides in length. For example the molecule may be about 5, 10, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 40, 45 or 50 nucleotides in length. Such polynucleotides may include from at least about 15 to more than about 120 nucleotides, including at least about 16 nucleotides, at least about 17 nucleotides, at least about 18 nucleotides, at least about 19 nucleotides, at least about 20 nucleotides, at least about 21 nucleotides, at least about 22 nucleotides, at least about 23 nucleotides, at least about 24 nucleotides, at least about 25 nucleotides, at least about 26 nucleotides, at least about 27 nucleotides, at least about 28 nucleotides, at least about 29 nucleotides, at least about 30 nucleotides, at least about 35 nucleotides, at least about 40 nucleotides, at least about 45 nucleotides, at least about 50 nucleotides, at least about 55 nucleotides, at least about 60 nucleotides, at least about 65 nucleotides, at least about 70 nucleotides, at least about 75 nucleotides, at least about 80 nucleotides, at least about 85 nucleotides, at least about 90 nucleotides, at least about 95 nucleotides, at least about 100 nucleotides, at least about 110 nucleotides, at least about 120 nucleotides or greater than 120 nucleotides.

[00075[The following example is provided to further illustrate the advantages and features of the present invention, but are not intended to limit the scope of the invention. While they are typical of those that might be used, other procedures, methodologies, or techniques known to those skilled in the art may alternatively be used.

EXAMPLE 1

PATERNAL SPERM DNA METHYLATION ANALYSIS

[00076[This example details a study in which genome-wide DNAm was examined in paternal semen biosamples obtained from an ASD enriched-risk pregnancy cohort, the Early Autism Risk Longitudinal Investigation (EARLI) cohort, to estimate associations between sperm DNAm and prospective ASD development, using a 12-month ASD symptoms assessment, the Autism Observation Scale for Infants (AOSI). Methylation data from 44 sperm samples were analyzed on the CHARM 3.0 array, which contains over 4 million probes (over 7 million CpG sites), including 30 samples also run on the Illumina Infinium HumanMethylation450™ (450k) BeadChip ™ platform (~485,000 CpG sites). Associated regions were also examined in an independent sample of postmortem human brain ASD and control samples for which Illumina 450K™ DNA methylation data were available.

[00077[EXPERIMENTAL MATERIALS AND METHODS

[00078[*Study sample*

[00079[Semen samples (n=44) were obtained from fathers of families enrolled in the Early Autism Risk Longitudinal Investigation (EARLI), an enriched autism risk pregnancy cohort that focuses on the prenatal and early life periods of newborns who have biological siblings already diagnosed with an ASD. For diseases such as ASD that demonstrate familial aggregation, disease-specific enriched-familial risk cohort studies represent an attractive alternative to large birth cohorts. In enriched-risk studies, subjects are over-sampled for increased familial risk, increasing power to detect associations between risk factors and disease. The EARLI study is comprised of four data collection sites (Drexel University, Johns Hopkins University, University of California, Davis, and Kaiser Permanente Division of Research) and includes prospective collection of biological samples from multiple family members along with environmental, questionnaire, interview, and clinical assessment data collected at multiple time points throughout the pregnancy and through age 3 of the newborn sibling. The EARLI study was reviewed and approved by Human Subjects Institutional Review Boards (IRBs) from each of the four study sites.

[00080[The samples analyzed within this project were collected from biological fathers of infant siblings using a home-collection semen kit distributed at the time of enrollment, typically in first or second trimester of pregnancy. Semen collection kits with instructions were mailed to the father of the current pregnancy prior to the first prenatal visit. The goal was to collect the semen sample as close to the time of the family's enrollment in EARLI as possible, but samples were accepted any time during the study. Paternal demographic information was collected via the EARLI paternal interview (or, if unavailable, from the maternal interview). Paternal age for each biosample was calculated by subtracting the father's date of birth from the date of sample collection.

[00081[A total of 44 sperm DNA samples were available for CHARM DNA methylation (DNAm) analysis, and an overlapping validation set of 30 samples with DNAm assessment via the Illumina 450k ™ platform (see methods below) was available. All 44 individuals had 12-month developmental assessment data from the child available.

[00082[*Phenotype assessment*

[00083[EARLI siblings participated in extensive neurodevelopmental phenotyping at regular intervals during development including clinician assessment and parent report. The 12-month clinical assessment included the Autism Observation Scale for Infants (AOSI) tool, an 18-item clinician-administered, semi-structured, play-based observation that is targeted for high-risk infants. It assesses eye contact, visual tracking, imitation, atypical sensory behavior, and social babbling (among other behaviors), many of which have been associated with later ASD. AOSI scores at 12-months have been shown to predict Autism Diagnostic Observation Schedule (ADOS) classification at 24- (9) and 36-months, and are thus used in this analysis as an early, quantitative, indicator of ASD related phenotype in at-risk siblings.

[00084[*Sample processing and DNA extraction*

[00085[Semen samples were frozen upon collection as per the EARLI collection kit instructions and then shipped (either same- or next-day) directly to the Johns Hopkins Biosample Repository (JHBR) for storage (-80°C) until processing. Genomic DNA for 44 samples used in CHARM analyses and for 26 of the samples used in 450K analyses was extracted from semen samples using a QIAgen QIAsymphony ™ automated workstation with the Blood 1000 ™ protocol of the DSP DNA Midi ™ kit (Cat. No. 937255, Qiagen, Valencia, CA) as per manufacturer's instructions. For the additional 4 samples measured via 450K analyses, DNA was extracted at the Johns Hopkins School of Medicine Center for Epigenetics using a version of the Epicentre MasterPure DNA Purification Kit ™ (Cat. No. MCD85201).

[00086[*CHARM DNA methylation measurement*

[00087[Genome-wide sperm DNA methylation was measured using the Comprehensive High-throughput Arrays for Relative Methylation assay (CHARM) (Ladd-Acosta et al., Comprehensive high-throughput arrays for relative methylation (CHARM). Current protocols in human genetics / editorial board, Jonathan L Haines et al. 2010 Apr; Chapter 20: Unit 20 1 1-19; incorporated herein by reference.). Genomic sperm DNA (4μg) was sheared on a HydroShear Plus™ (DigiLab, Marlborough, MA), digested with McrBC, gel-purified, labeled, and hybridized to arrays as described (Ladd-Acosta et al., supra). The CHARM 3.0™ array (Roche NimbleGen™, Madison, WI) includes over 4 million probes and covers over 7 million CpGs arranged into probe groups (consecutive probes are within 300 bp of each other). These arrays include probes covering all annotated and non-annotated promoters and microRNA sites in addition to features present in the original CHARM method (Ladd-Acosta et al., supra).

[00088[*Illumina 450k™ DNA methylation measurement*

[00089[DNAm for 30 overlapping paternal semen samples was measured via the Illumina Infinium HumanMethylation450 BeadChip ™ assay (Illumina, San Diego, CA). Genomic DNA (1μg) was processed by the Johns Hopkins University SNP Center using the automated Infinium™ workflow. Technical control samples, including technical replicates of liver and placenta, and spike-in samples (here, 0%, 50%, and 100% methylated commercial DNA) were repeated across plates to ensure consistent high quality data. Note these samples were not included in any subsequent analyses of semen.

[00090[*Data pre-processing*

[00091[CHARM raw data were pre-processed using the CHARM Package™ (v.2.8.0) in R™ (version 3.0.3). Briefly, probe-level percentage DNAm estimates were obtained by first removing background signal using anti-genomic background probes, followed by normalization using control probes (loess for within arrays, followed by quantile between arrays). After normalization, we excluded background, control and repetitive probe groups, yielding 3,811,046 total probes per array for each of the 44 discovery samples. Surrogate variable analysis (SVA) was then performed on these percentage methylation estimates to estimate latent factors influencing DNAm levels which typically represent "batch" effects. The number of surrogate variables (SVs) to include in the statistical models were estimated using the Buja and Eyuboglu (“be”) algorithm, which identifies how many latent variables are present in the data. Then the SVA algorithm estimates these SVs which are adjusted for as confounders in downstream differential methylation bump hunting analysis.

[00092[Validation 450k raw data were pre-processed using the Minfi™ Package (v.1.10.1) in R™ (version 3.1.0). Data were stratified quantile normalized and SVA was performed on the percentage methylation estimates (beta scale) to adjust for potential batch effects.

[00093[*Age and developmental assessment associations*

[00094[Differences in AOSI score across demographic and technical/laboratory variables were compared using F-tests. Bivariate associations between 12-month AOSI scores and ordinal variables were calculated using linear model trend tests. Similar tests were also performed to assess the degree of association between SVs and demographic/laboratory variables. Spearman rank correlations were estimated between paternal age and offspring 12-month AOSI scores.

[00095[*Regional DNAm discovery*

[00096[Regions of CHARM DNAm differences by infant AOSI score were identified using the “bump hunting” approach previously developed for the CHARM platform (Jaffe et al., 2012 Feb;41(1):200-9), adjusting for estimated surrogate variables. The statistical model treated AOSI as the outcome of interest, and adjusted for paternal age and 10 surrogate variables (described above) as confounders. This model is applied to every high-quality probe on the microarray to identify the adjusted linear effect of AOSI on DNA methylation levels. Regions of differential methylation were identified by smoothing the linear effects of AOSI on DNA methylation, and then thresholding these smoothed statistics across all probes using the 99.9th percentile as a cutoff, as previously described (Jaffe et al., supra). P-values for each DMR were calculated from a genome-wide empirical distribution of null statistics generated using a linear model bootstrapping approach across 10,000 permutations as described (Jaffe et al., supra). DMRs with a genome-wide family-wise empirical p value (FWER) < 0.05 were identified. Note that the DMRs are ranked prior to the permutation procedure, which is used to identify the threshold that controls for the target genome-wide FWER. Mean methylation values across the DMR for each sample were also plotted by AOSI score and Spearman rank correlation coefficients estimated.

[00097[Post-hoc sensitivity analyses was performed to assess the potential influence of skewed or extreme AOSI scores and the influence of race on identified DMRs. First, to account for right-skewed AOSI scores, a natural log(AOSI+1) transformation was used in regressions against the average methylation per DMR, for each DMR with $p < 0.05$. A dichotomous outcome was also considered comparing the highest quartile of AOSI scores (>10) to the lowest three quartiles. These provide comparisons of effect size magnitude and direction between the discovery DMRs and these alternative forms of AOSI modeling. Bump hunting models were also reanalyzed adjusting for principal components of race (PC1 and PC2 from ancestry analysis of GWAS SNP data) and beta coefficients compared between this race-adjusted model and the primary discovery bump hunting results.

[00098[*Gene enrichment analyses*

[00099[Enrichment of genes was tested for within 10kb of DMRs with FWER < 0.05 based on Gene Ontology^{TM} (Biological Processes database) using the hypergeometric test restricted to gene sets with at least 4 members. The GOstats R Bioconductor^{TM} package was used to compare genes mapped to top DMRs (FWER < 0.05) to all genes on the CHARM array that also have an Entrez ID as background.

[000100[*Cross-platform validation*

[000101[DMRs discovered via CHARM for AOSI score were attempted to be validated using overlapping genomic coverage on the 450k array in a partially overlapping set of paternal samples. Within these subsets of genomic sites, linear regression was first used to test the relationship between single-site 450k DNAm and AOSI score. Statistical models were adjusted for surrogate variables estimated from the data of only overlapping samples (n=30 samples with 7 SVs for AOSI score). Spearman correlation tests were used to calculate the correlation between effect estimates from CHARM and 450k.

[000102[*Comparison with autism brain data*

[000103[Publicly available Illumina 450k^{TM} data from postmortem human brain tissue was downloaded on 40 samples (19 ASD and 21 controls) across 3 brain regions (frontal cortex, temporal cortex, and cerebellum) from GEO dataset: GSE53162 (25). Illumina 450k^{TM} data were normalized as above with the FlowSorted.DLPFC.450k Bioconductor^{TM} dataset to estimate cellular composition in each sample. Mean differences and resulting t-statistics and p-values were calculated within each brain region comparing cases to controls using the limma Bioconductor^{TM} package, and considered probes with differential methylation with $p < 0.05$ marginally significant for the cross-tissue comparison analysis.

[000104[RESULTS

[000105[*Study sample characteristics*

[000106[Infant sibling performance on the 12-month AOSI for children of the fathers examined in the CHARM discovery sample ranged from 0 to 18. Fathers were predominantly white and non-Hispanic (73%) and their ages ranged between 27 to 51 years (Table 1). Sibling AOSI scores were not associated with paternal age (ρ= -0.04), education level or family income, though higher AOSI scores were observed at the Drexel study site.

[000107[*Methylation data quality assessment*

[000108[Total AOSI did not vary by CHARM DNA shearing date (*P* value=0.77), HydroShear™ machine (*P* value=0.9) CHARM gel (*P* value=0.65), sample location in CHARM gel (*P* value=0.71), or CHARM hybridization date (*P* value=0.39). Table 4 shows the degree of association between CHARM variables to each of the estimated surrogate variables.

[000109[Illumina 450k™ between-plate correlation Pearson coefficients of unnormalized DNAm between technical control replicates ranged from 0.848 to 0.996 with a mean (SD) of 0.963 (0.041). DNA samples were hybridized to Illumina arrays across 3 dates and AOSI score did vary by hybridization date (*P* value=0.06). Table 5 shows the degree of association between 450k variables to each of the estimated surrogate variables.

[000110[*DNAm and AOSI score*

[000111[2605 candidate DMRs were identified using our "bump hunting" method in paternal sperm associated with child AOSI score at 12 months, and after permutation analysis, the top 193 DMRs had genome-wide $p < 0.05$. The top 10 ranked DMRs are shown in Table 2, and Table 6 includes the complete list of 193 DMRs. Methylation plots by AOSI score for all 193 DMRs are publicly available on the World Wide Web at URL ije.oxfordjournals.org/content/44/4/1199.full. The top four regions where DNAm associates linearly (genome-wide p=0.001) with AOSI score were highlighted in Figure 1, with the average DNAm across the DMR shown in each inset. Figure 1a shows hypomethylation with increased AOSI score on chromosome 15 overlapping *SNORD115-15* where the highest quartile of AOSI scores ($\geq$10) corresponded to 27.4% average sperm methylation compared with 46.9% average methylation for the lowest quartile (< 3). Similarly, Figures 1b and 1c show hypomethylation with increased AOSI scores on chromosome 15 covering *SNORD115-11* and *SNORD115-17*, respectively, with differences in regional average methylation between the highest and lowest AOSI quartiles of 19.7% and 22.6%, respectively. Figure 1d illustrates hypermethylation with increased AOSI scores on chromosome 1 overlapping *SMYD3*, with 68.9% average methylation for the highest AOSI score quartile compared to 47.3% average methylation of the lowest quartile.

[000112[An inverse relationship between AOSI and DNAm was observed at 141 (73%) of these regions. These DMRs overlapped genes (87 were located within 10 kb of a gene) that were enriched for gene ontology biological processes of cellular movement, neurogenesis, and neuronal development categories (Table 3). Sensitivity analysis involving linear regression on log-transformed AOSI scores [ln(AOSI+1)] yielded consistent results with the bump hunting performed on the linear AOSI models described above – all 193 DMRs matched directionality with regression statistics and 144/193 regions (74.6%) had $p < 0.05$. Similarly, when performing linear regression using a model of the highest quartile of AOSI score relative to the other three quartiles, the results were also consistent with the primary analyses – all 193 DMRs matched directionality with regression statistics and 142/193 regions (73.6%) had $p < 0.05$. Results from bump hunting analyses that adjusted for principal components of race were not substantially different from those reported. Beta coefficients for DMRs before and after race adjustment were highly correlated ($\rho = 0.99$, Figure 3).

[000113[Among these 193 AOSI-associated DMRs, Illumina 450k™ probes located within 500bp of the region boundaries were extracted. This was possible for 75 (38.9%) CHARM-discovered DMRs (Table 7) represented by 244 probes on the 450K array. The direction of association between AOSI and DNAm was consistent with CHARM results for 74 (98.7%) of these 75 regions (Figure 2).

[000114[*AOSI-associated DMRs in postmortem brain*

[000115[A subset of the AOSI-associated DMRs identified in semen also shows directionally consistent effects in postmortem human brain samples in a study comparing patients with ASD to unaffected controls. The AOSI DMRs were most consistent comparing cases and controls in the cerebellum, where 21 of 75 of AOSI-associated DMRs contained at least one Illumina 450K probe that was differentially methylated comparing autism to controls; 18 of these were directionally consistent with the sperm AOSI score association data (*e.g.* DNAm positively associated with both AOSI score in semen and in ASD brains). This cross-tissue consistency was not present in the cortical brain regions – in the frontal cortex, only 12 AOSI DMRs were significant, of which 3 were directionally consistent, and in the temporal cortex, only 10 AOSI DMRs were significant, of which 2 were directionally consistent. Surprisingly, 15 of 18 cerebellum-consistent AOSI DMRs were overlapping genes in the SNORD families (Table 8).

[000116[FIGURE LEGENDS

[000117[Figures 1A-1D are graphical representations of methylation plots for the top 4 statistical DMRs identified using CHARM and 12-month score in embodiments of the present invention. (A) SNORD115-15, (B) SNORD115-15, (C) SNORD115-17, (D) SMYD3. Top panels show individual methylation levels at each probe by genomic position. Dotted vertical lines represent the boundaries of the DMR, and coloured lines represent the average methylation curve for samples grouped by quartiles of AOSI scores—the scores within each quartile are shown in the legend. Middle panel shows location of CpG dinucleotides (as black tick marks) and CpG density by genomic position (black curved line). Bottom panel shows location (boxes) and direction (+ or −) of refseq-annotated genes. Inset scatterplot reflects linear regression of average methylation across all probes within DMR per sample by AOSI 12-month score.

[000118[Figure 2 is a graphical representation depicting validation of CHARM AOSI DMR values using 450k data from overlapping samples (n=30). The CHARM DMR value corresponds to the smoothed effect estimate at each probe, plotted against the single-site regression coefficients from 450 K data for each corresponding DMR. Spearman rank correlation coefficient is also included.

[000119[Figure 3 is a graphical representation showing the relationship between linear regression beta coefficients of the original bump hunting model (Beta1) and the model including principal components of race from ancestry analysis of GWAS SNP data (Beta2). Spearman rank correlation coefficient also included.

[000120[Figure 4 sets forth Table 1 relating to bivariate associations of AOSI score with demographic and laboratory variables for the EARLI study population.

[000121[Figure 5 sets forth Table 2 relating to candidate differentially methylated regions (DMRs) associated with 12-month total AOSI score, an indicator of autism risk.

[000122[Figure 6 sets forth Table 3 relating to gene ontology, biological processes enriched in differentially methylated regions associated with 12-month AOSI score.

[000123[Figure 7 sets forth Table 4 relating to association between SVs and variables for CHARM samples.

[000124[Figure 8 sets forth Table 5 relating to association between SVs and variables for 450k samples.

[000125[Figure 9 sets forth Table 6 relating to the top 193 DMRs associated with autism as determined by an embodiment of the invention.

[000126[Figure 10 sets forth Table 7 relating to validation of 450k probes.

[000127[Figure 11 sets forth Table 8 relating to overlap of AOSI DMRs.

[000128[DISCUSSION

[000129[In summary, a strong relationship between DNA methylation in sperm from fathers and 12-month ASD-related phenotype in infant siblings within an enriched autism risk birth cohort was observed. The degree of methylation change was quite substantial compared to recent reports for other epigenome-wide association studies (EWAS): 6.2 – 30.7% difference between top and bottom quartiles of AOSI score among the 193 DMRs with FWER < 0.05. These AOSI-associated DMRs showed enrichment for genes involved in neurogenesis and more general neuronal development.

[000130[12-month infant developmental assessment as a measure of early ASD-related phenotype was focused on. Total scores on the AOSI administered to high-risk infants at 12-months has shown high inter-rater reliability (0.93), test-retest reliability (0.61) and correlation with ASD diagnosis at age three. Because of the unusual nature of this prospective high-risk cohort, direct replication is difficult. However, an independent data set of autism and control brains showed overlap with the AOSI-associated DMRs, with 18 of 21 genes directionally consistent in methylation values.

[000131[How might these DMRs contribute to ASD? The sperm contributes half of the genome to somatic cells, and after germline reprogramming (except for generating germ cells in the next generation), this would likely persist into brain development. The fact that these genes were enriched for neurological function is consistent with this idea. Remarkably, 15 members of the small nucleolar RNA, C/D Box (SNORD-115) genes were in this overlapping set. This gene cluster lies within the Prader-Willi critical region of chromosome 15 and are considered candidate mechanistic targets for the disorder, a canonical imprinted Mendelian disorder with autism features. Gene dosage alterations in this region are also associated with ASD. The data reported here suggest that paternal epigenetic changes of this gene cluster confer risk of ASD among offspring, at least among those with an older affected sibling. Molecularly, aberrant imprinting, as seen in Prader-Willi syndrome, would not explain these results as the paternal allele is normally expressed.

[000132[The association with the cerebellum is intriguing. A case report of twins discordant for ASD showed cerebellar anatomical changes in the affected twin, and some neuro-imaging studies have shown changes in cerebellum in autism, or autistic features in children with cerebellar defects.

[000133[One potential limitation of the study is that DNA was extracted from semen rather than purified sperm. However DNA methylation has been shown to be highly uniform regardless the quality of sperm subpopulations. Thus, there was no adjustment for sperm quality or subpopulations. Also, there has been recent increased awareness of possible bias in enrichment statistics when gene size is not considered. No difference between gene size ($p=0.055$), coding length ($p=0.79$), and GC content ($p=0.1$) comparing DMRs with FWER < 0.05 ($N=193$) to those with FWER > 0.05 ($N=2412$) was found. Further, the average gene length was actually shorter for AOSI DMRs in our results (144kb vs 178kb). Additionally, the CHARM design features sets of probe groups, each with very evenly spaced probes (56-90bp), which was designed solely to target regions of at least moderate CpG density in the genome in an annotation-agnostic manner. Note this is different than the more promoter-focused designs, where probe density has been shown to bias gene set analysis. The CHARM method is less quantitative at individual loci due to the use of a restriction enzyme and how signal is measured. However, it was previously shown that it is more powerful for identifying relative changes in DNAm at the region level.

[000134[While the prevalence of autism spectrum disorder (ASD) has been increasing in recent years, even the current estimate of 1-2% (2) makes study of a general population prospective pregnancy cohort not feasible for ASD. For diseases such as ASD that demonstrate familial aggregation, disease-specific enriched-familial risk cohort studies are a commonly employed design. There is a long history of enriched-risk (family-based) cohort studies in epidemiology to examine risk factors for rare diseases. These studies are not designed to be generalizable to the general population, though targeted results from an enriched risk cohort can be replicated in a representative population to test generalizability. Nevertheless, in our study, fathers of second children with high- and low- AOSI scores are from the same underlying EARLI enriched-risk population, and thus effect estimates of association are not distorted due to selection bias, but in fact internally valid.

[000135[In summary, relatively large inter-individual differences in paternal sperm DNAm were identified that associate with later 12-month ASD-related phenotype in their biological offspring, among a sample of enriched-risk ASD infant siblings. Many of these regions of paternal sperm DNAm were also associated with ASD in cerebellum brain samples. Further, the genes implicated are enriched for neurodevelopment and include regions implicated in Prader-Willi syndrome. This unique set of biosamples is comparatively small, given the nature of the biosample type and required perinatal collection timing in an enriched risk cohort. Thus it is important to examine these findings in larger samples to confirm these associations and explore mechanistic implications.

[000136[Although the invention has been described with reference to the above example, it will be understood that modifications and variations are encompassed within the spirit and scope of the invention. Accordingly, the invention is limited only by the following claims.

What is claimed is:

1. A method for determining risk of autism spectrum disorder (ASD) in an offspring subject comprising:

a) analyzing DNA methylation status in a sample containing sperm from the prospective paternal parent; and

b) determining a deviation in the methylation status of (a) as compared to a corresponding normal sample not associated with ASD, wherein a methylation pattern that deviates from the pattern found in the corresponding sample not associated with ASD, is indicative of a risk of ASD in the offspring.

2. A method for determining whether a subject has or is at risk of having autism spectrum disorder (ASD) comprising:

a) analyzing DNA methylation status in a DNA sample of the subject; and

b) determining a deviation in the methylation status of (a) as compared to a corresponding normal sample not associated with ASD, wherein a methylation pattern that is different from the pattern found in the corresponding sample not associated with ASD, is indicative of a subject having or being at risk of having ASD.

3. The method of claims 1 or 2, wherein the subject is a human.

4. The method of claim 2, wherein the sample is a blood or tissue sample.

5. The method of claim 1, wherein the sample is a semen sample.

6. The method of claims 1 or 2, wherein determining a methylation status comprises determining the methylation at differentially methylated regions (DMRs) in the DNA.

7. The method of claims 1 or 2, wherein the methylation status is performed by one or more techniques selected from the group consisting of a nucleic acid amplification, polymerase chain reaction (PCR), methylation specific PCR, bisulfite sequencing, capture bisulfite sequencing, whole genome bisulfite sequencing, pyrosequencing, single-strand conformation polymorphism (SSCP) analysis, restriction analysis, microarray technology, including bead microarray technology, and proteomics.

8. The method of claims 1 or 2, comprising performing comprehensive high-through array-based relative methylation (CHARM) analysis on a sample of labeled, digested genomic DNA.

9. The method of claims 1 or 2, comprising performing methylation analysis via a bead array.

10. The method of claim 6, wherein the DMRs are associated with genes for neurogenesis and/or neuronal development.

11. The method of claim 6, wherein the DMRs are selected from the DMRs as set forth in Table 2, Table, 6, Table 7, or Table 8.

12. The method of claim 11, wherein the DMRs are selected from the DMRs as set forth in Table 2.

13. The method of claim 6, wherein the DMRs are associated with a gene ontology profile set forth in Table 3.

14. The method of claim 1, wherein methylation status is determined for a gene, or regulatory region thereof, set forth in Table 2, Table, 6, Table 7, or Table 8.

15. The method of claim 1 or 2, further comprising providing assessment of risk or developing ASD.

16. The method of claim 15, wherein the risk is assessed as a score relative to low, moderate or high risk.

17. The method of claim 10, wherein the DMR is associated with chromosome 15.

18. The method of claim 17, wherein the DMR is associated with Prader-Willi syndrome.

19. The method of claims 1 or 2, wherein the methylation status is hypomethylated or hypermethylated.

20. The method of claim 11, wherein methylation status is determined for at least 10 DMRs selected from those set forth in Table 2, Table, 6, Table 7, and Table 8.

21. The method of claim 11, wherein methylation status is determined for at least 25 DMRs selected from those set forth in Table 2, Table, 6, Table 7, and Table 8.

22. The method of claim 11, wherein methylation status is determined for at least 50 DMRs selected from those set forth in Table 2, Table, 6, Table 7, and Table 8.

23. A plurality of nucleic acid sequences that selectively hybridizing to a nucleic acid sequence selected from the group consisting of differentially methylated regions (DMRs) set forth in Table 2, Table, 6, Table 7, and Table 8.

24. The plurality of nucleic acid sequences of claim 23, wherein the plurality comprises nucleic acid sequences that selectively hybridize to at least 10 DMRs selected from those set forth Table 2, Table, 6, Table 7, and Table 8.

25. The plurality of nucleic acid sequences of claim 23, wherein the plurality comprises nucleic acid sequences that selectively hybridize to at least 25 DMRs selected from those set forth Table 2, Table, 6, Table 7, and Table 8.

26. The plurality of nucleic acid sequences of claim 23, wherein the plurality comprises nucleic acid sequences that selectively hybridize to at least 50 DMRs selected from those set forth Table 2, Table, 6, Table 7, and Table 8.

27. The plurality of nucleic acid sequences of claim 23, wherein the plurality comprises nucleic acid sequences that selectively hybridize to the DMRs set forth in Table 2.

28. The plurality of nucleic acid sequences of claim 23, wherein each nucleic acid sequence is about 10-55 base pairs in length.

29. The plurality of nucleic acid sequences of claim 23, wherein each nucleic acid sequence is about 25-35 base pairs in length.

30. The plurality of nucleic acid sequences of claim 23, wherein each nucleic acid sequence is about 35-45 base pairs in length.

31. The plurality of nucleic acid sequences of claim 23, wherein each nucleic acid sequence is about 45-55 base pairs in length.

32. A microarray comprising the plurality of sequences claim 23.

33. A method of determining methylation status of genomic DNA isolated from a cell comprising:

performing an assay utilizing the microarray of claim 32.

34. The method of claim 33, wherein the assay is comprehensive high-throughput array for relative methylation (CHARM) analysis on a sample of labeled, digested genomic DNA isolated from the cell.

35. The method of claim 33, wherein the assay utilizes a bead array format.

36. The method of claim 33, further comprising:

a) determining the methylation status of at least 10 DMRs selected from those set forth in Table 2, Table, 6, Table 7, and Table 8.; and

b) identifying or classifying risk of ASD.

37. The method of claim 36, wherein methylation status is determined for at least 25 DMRs selected from those set forth Table 2, Table, 6, Table 7, and Table 8.

38. The method of claim 36, wherein methylation status is determined for at least 50 DMRs selected from those set forth Table 2, Table, 6, Table 7, and Table 8.

39. A kit for determining whether a subject has or is at risk of having or inheriting a risk of having ASD, comprising:

a) a reagent for determining methylation status of a nucleic acid sequence, wherein the nucleic acid sequence is a differentially methylated region (DMR) associated with ASD; and

b) instructions for use of the reagent.

40. The kit of claim 39, wherein the reagent is an oligonucleotide probe, primer, or primer pair, or combination thereof, capable of selectively hybridizing to the DMR, with or without prior bisulfite treatment of the DMR.

41. The kit of claim 39 or 40, further comprising one or more detectable labels.

42. The kit of claim 39, wherein the DMR is selected from those set forth Table 2, Table, 6, Table 7, and Table 8.

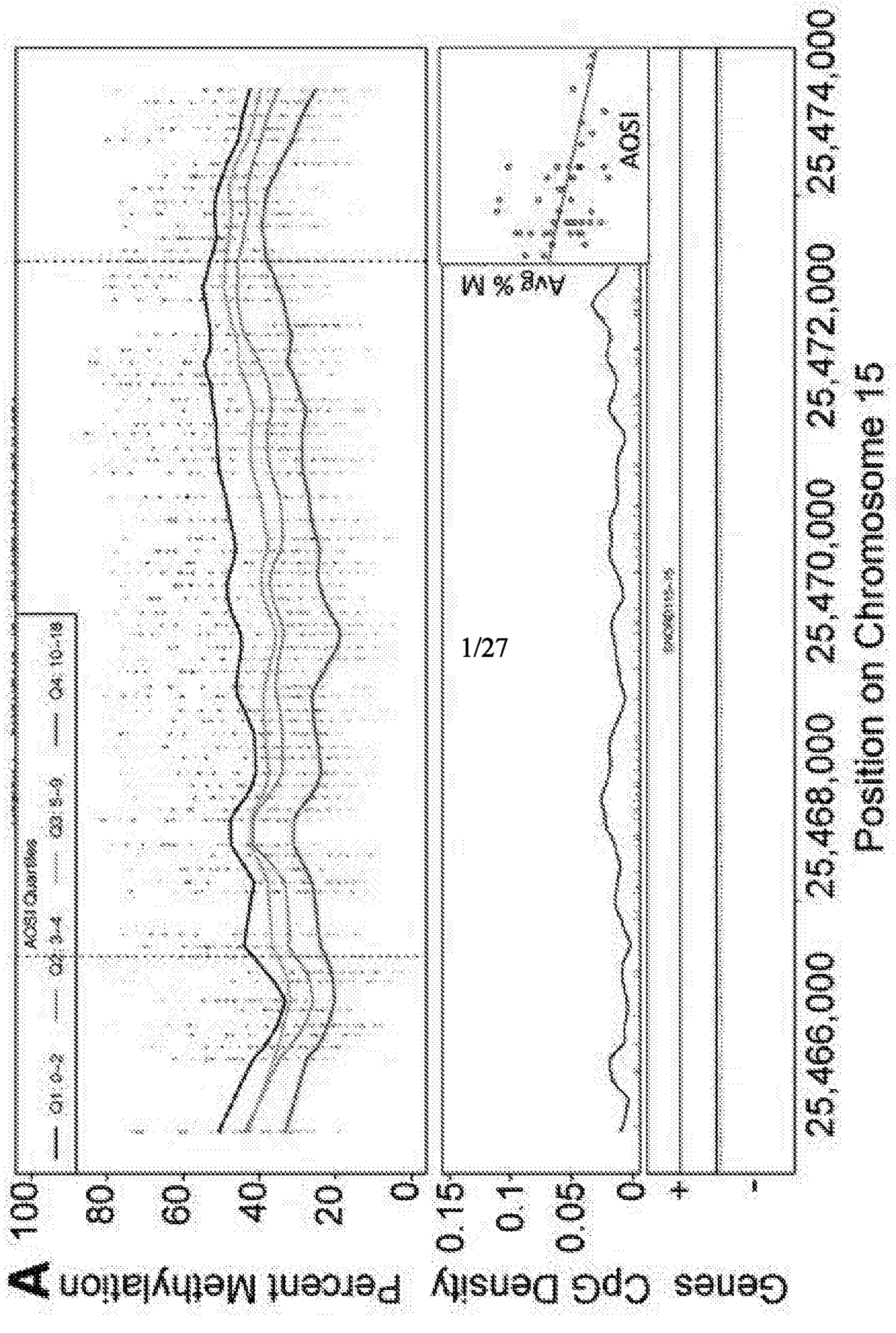
A
Percent Methylation
CpG Density
Genes
100
80
60
40
20
0
0.15
0.1
0.05
0
+
-
Avg % M
AOSI
1/27
25,466,000
25,468,000
25,470,000
25,472,000
25,474,000
Position on Chromosome 15

FIG. 1A

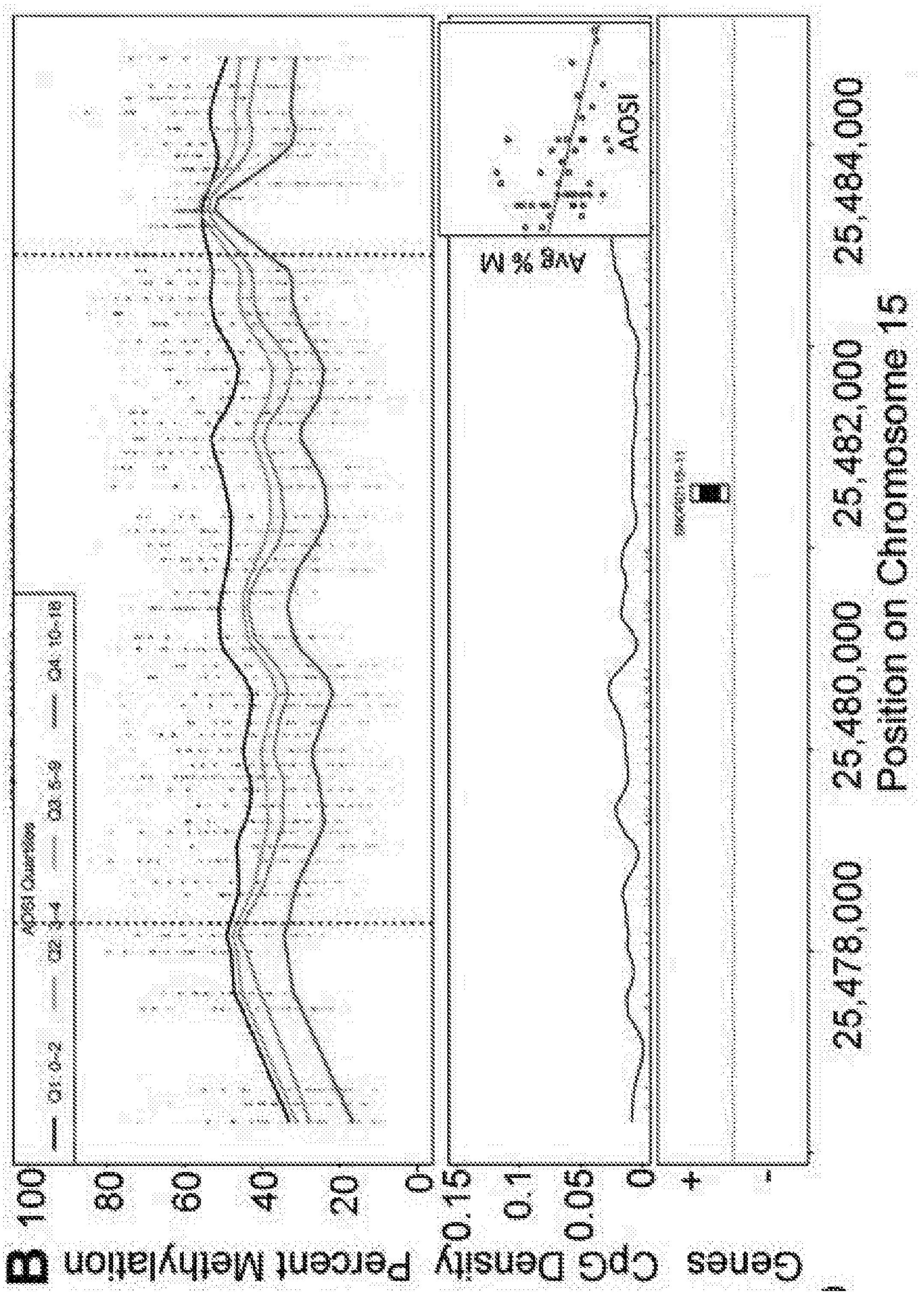
B
Percent Methylation
CpG Density
Genes
100
80
60
40
20
0
0.15
0.1
0.05
0
+
-
Q1: 0-2
Q2: 3-4
Q3: 5-9
Q4: 10-18
Avg % M
AOSI
25,478,000
25,480,000
25,482,000
25,484,000
Position on Chromosome 15

FIG. 1B

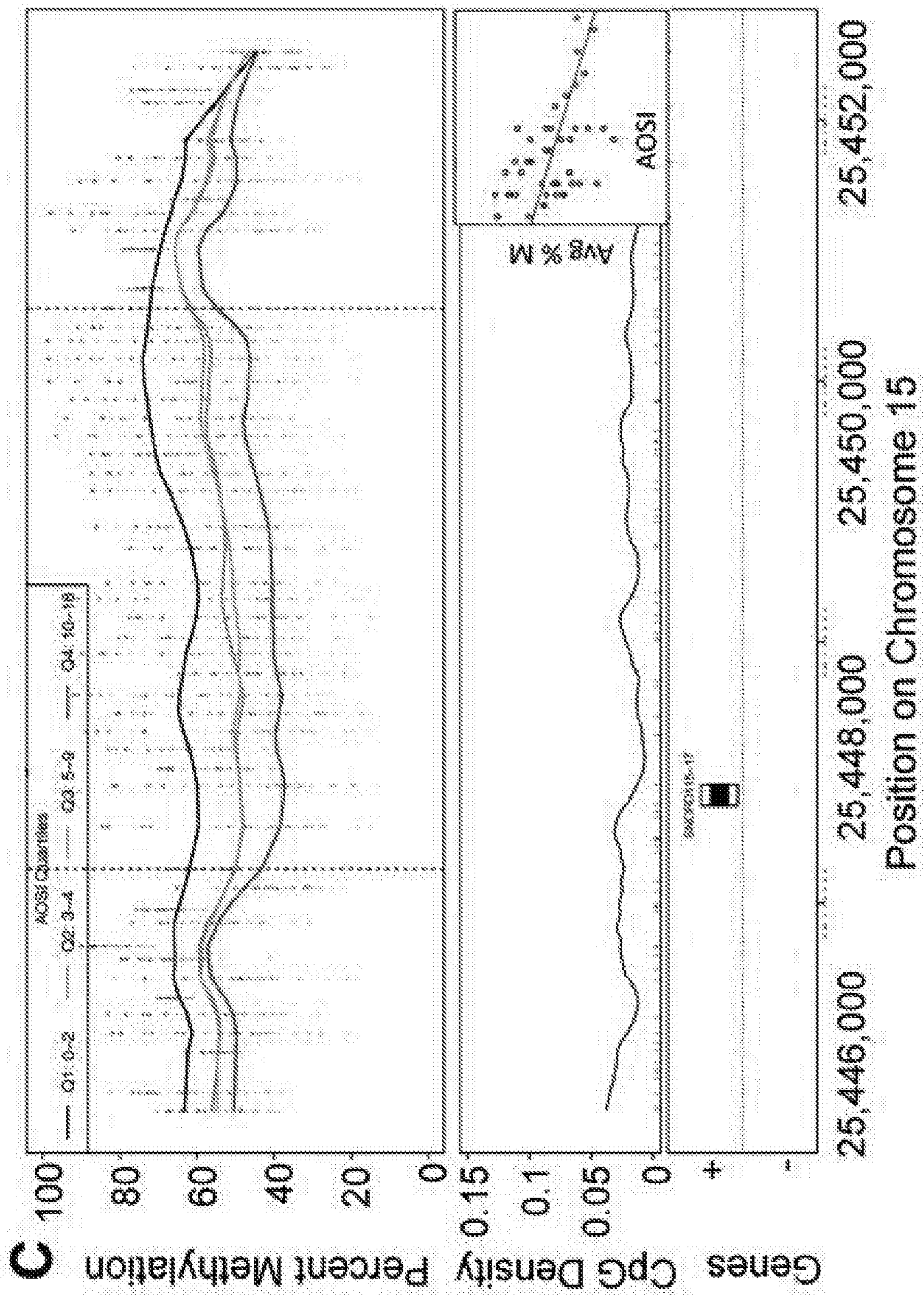
C
Percent Methylation
CpG Density
Genes
100
80
60
40
20
0
0.15
0.1
0.05
0
+
-
AOSI Quartiles
Q1: 0~2
Q2: 3~4
Q3: 5~9
Q4: 10~19
Avg % M
AOSI
25,446,000
25,448,000
25,450,000
25,452,000
Position on Chromosome 15

FIG. 1C

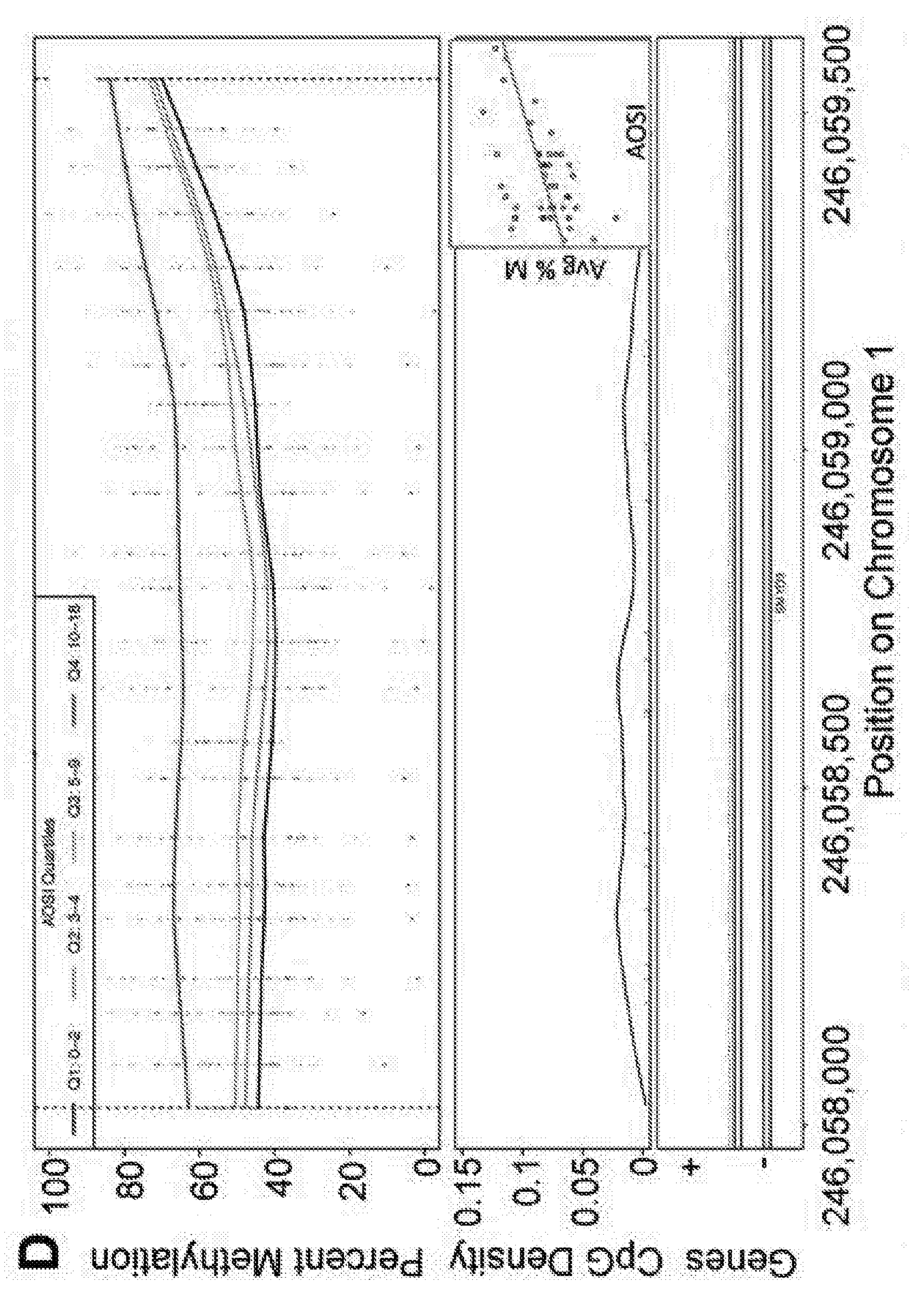
D
Percent Methylation
CpG Density
Genes
100
80
60
40
20
0
0.15
0.1
0.05
0
+
-
AOSI Quartiles
Q1: 0–2
Q2: 3–4
Q3: 5–8
Q4: 10–18
Avg % M
AOSI
246,058,000
246,058,500
246,059,000
246,059,500
Position on Chromosome 1

FIG. 1D

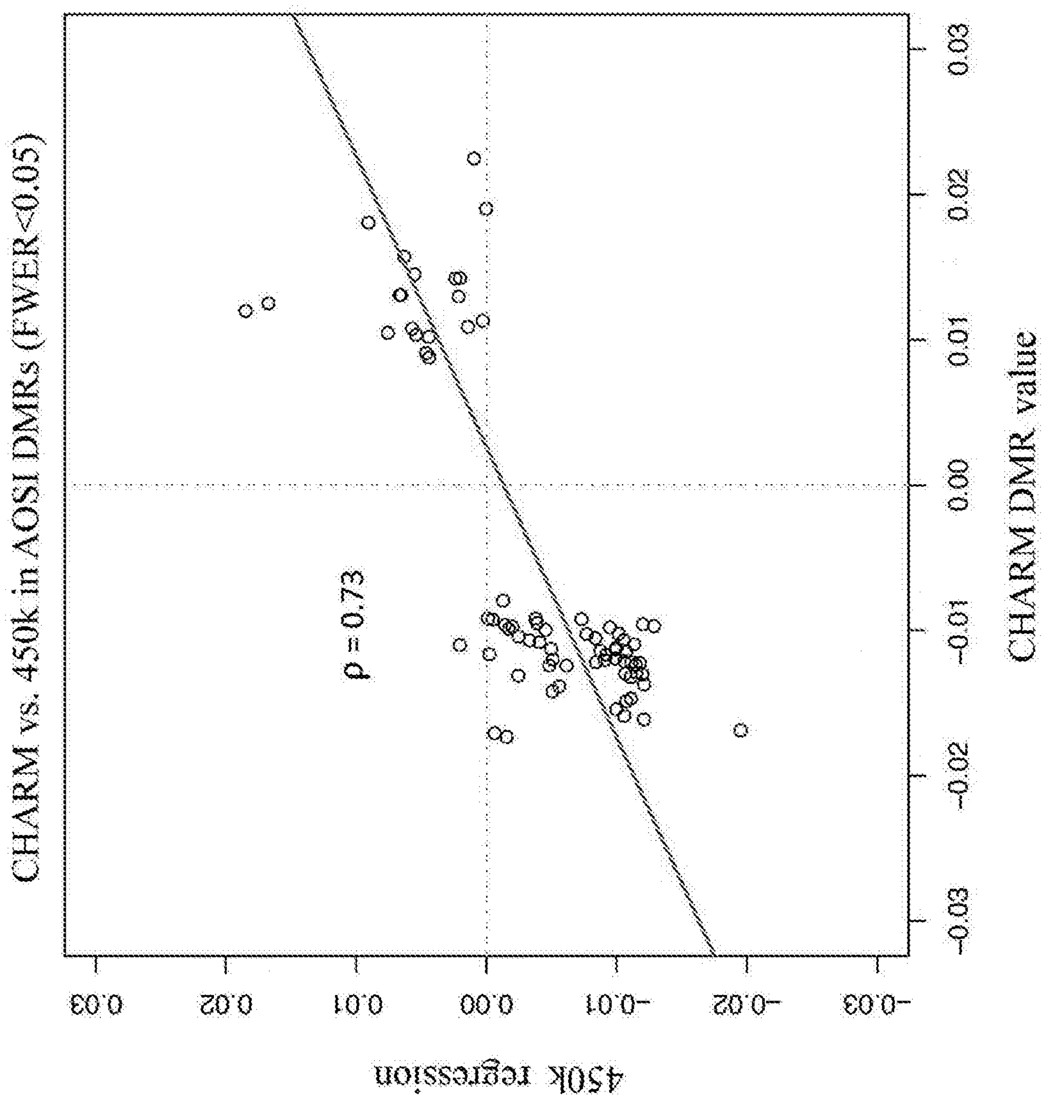
CHARM vs. 450k in AOSI DMRs (FWER<0.05)
ρ = 0.73
CHARM DMR value
450k regression
−0.03
−0.02
−0.01
0.00
0.01
0.02
0.03

FIG. 2

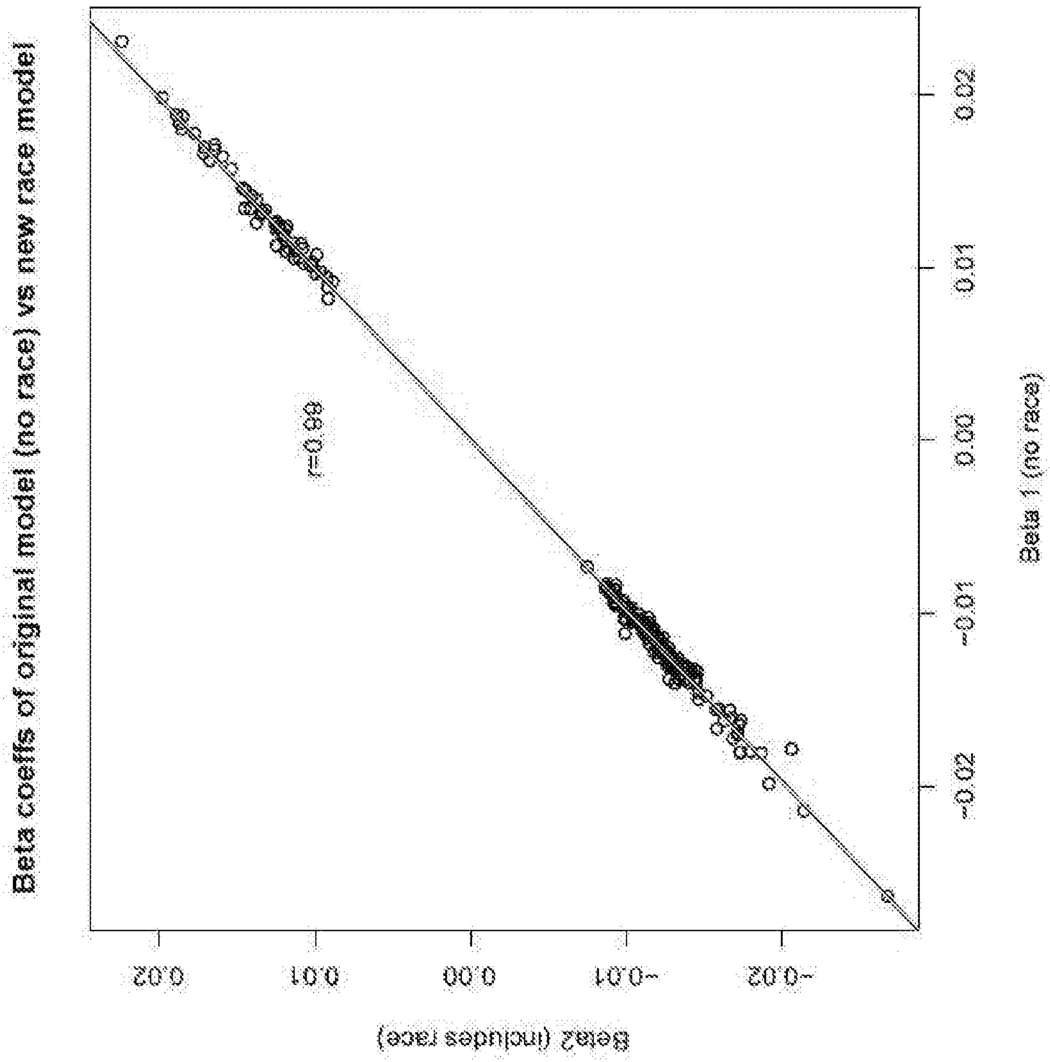
Beta coeffs of original model (no race) vs new race model
r=0.99
Beta2 (includes race)
Beta 1 (no race)
0.02
0.01
0.00
-0.01
-0.02

FIG. 3

Table 1: Bivariate associations of AOSI score with demographic and laboratory variables for the EARLI study population

Total 12-month AOSI score ($N = 44$)		CHARM ($N = 44$)	Subset also with 450 K ($N = 30$)	
Variable	**Mean (SD)**	***P*-value**	***N* (%)**	***N* (%)**
Age quartile				
27–32.7	6.5 (4.5)		12 (27.3)	9 (30)
32.8–36	5 (5.8)		10 (22.7)	7 (23.3)
36.1–40.3	5.7 (4)		11 (25)	10 (33.3)
40.2–51.2	5.4 (3.7)	0.88	11 (25)	4 (13.3)
Race, ethnicity				
Black, Hispanic	7 (NA)		1 (2.3)	1 (3.3)
Black, non-Hispanic	4.7 (2.5)		3 (6.8)	2 (6.7)
Other, Hispanic	7.5 (0.7)		2 (4.5)	1 (3.3)
Other, non- Hispanic	5 (4.2)		2 (4.5)	1 (3.3)
White, Hispanic	3 (NA)		1 (2.3)	1 (3.3)
White, non- Hispanic	5.4 (4.5)		32 (72.7)	22 (73.3)
White, other	12 (8.5)	0.63	2 (4.5)	2 (6.7)
Missing	4 (NA)		1 (2.3)	-
Study site				
Drexel	8.5 (5.5)		14 (31.8)	10 (33.3)
Johns Hopkins	3.6 (2.8)		15 (34.1)	9 (30)
Kaiser	4.3 (3.8)		7 (15.9)	6 (20)
UC Davis	5.9 (2.4)	0.014	8 (18.2)	5 (16.7)
Paternity status				
Proband and sibling	5.6 (4.6)		37 (84.1)	24 (80)
Sibling only	5.9 (3.2)	0.91	7 (15.9)	6 (20)
Paternal education				
High school graduate or less	5.5 (2.1)		8 (18.2)	5 (16.7)
Some college	8.2 (6.2)		9 (20.5)	6 (20)
Bachelor's degree	4.8 (3.8)		13 (29.5)	9 (30)

FIG. 4

Total 12-month AOSI score ($N = 44$)		CHARM ($N = 44$)	Subset also with 450 K ($N = 30$)	
Variable	**Mean (SD)**	***P*-value**	***N* (%)**	***N* (%)**
Graduate degree	4.9 (4.3)	0.27	14 (31.8)	10 (33.3)
Annual income				
< US$49,999	5.6 (2.3)		9 (20.5)	6 (20)
$50,000–$74,999	7.6 (7.2)		5 (11.4)	4 (13.3)
$75,000–$99,999	5.5 (4.8)		13 (29.5)	10 (33.3)
>=$100,000	5.4 (4.2)	0.79	17 (38.6)	10 (33.3)
Shear date				
1	5.9 (4.5)		9 (20.5)	-
2	5.1 (4.4)		18 (40.9)	-
3	6.2 (4.4)	0.77	17 (38.6)	-
Shear machine (HS)				
2	5.6 (5.2)		19 (43.2)	-
3	5.9 (3.8)		22 (50)	-
4	4.7 (4.2)	0.90	3 (6.8)	-
CHARM gel				
1	0 (NA)		1 (2.3)	-
2	5.2 (3.2)		4 (9.1)	-
3	8 (5.1)		4 (9.1)	-
4	7 (7.4)		4 (9.1)	-
5	3.3 (1.5)		3 (6.8)	-
6	2 (0)		2 (4.5)	-
7	4.2 (3.9)		4 (9.1)	-
8	8 (4)		3 (6.8)	-
9	4.5 (3.5)		2 (4.5)	-
10	8.3 (8.5)		3 (6.8)	-
11	7.2 (4.9)		4 (9.1)	-
12	3 (1)		3 (6.8)	-
13	7.8 (2.2)		4 (9.1)	-

FIG. 4 (Con't.)

Total 12-month AOSI score ($N = 44$)		CHARM ($N = 44$)	Subset also with 450 K ($N = 30$)	
Variable	**Mean (SD)**	***P*-value**	***N* (%)**	***N* (%)**
14	3.7 (1.5)	0.65	3 (6.8)	-
Gel location				
Bottom middle	2 (NA)		1 (2.3)	-
Bottom left	4.6 (4.5)		12 (27.3)	-
Bottom right	6.9 (2.9)		8 (18.2)	-
Top left	5.8 (5.6)		12 (27.3)	-
Top right	6.2 (3.8)	0.71	11 (25)	-
Hybridization date				
1	5 (4.5)		20 (45.5)	20 (66.7)
2	6.2 (4.3)	0.39	24 (54.5)	8 (26.7)
3	-		-	2 (6.7)

FIG. 4 (Con't.)

Table 2: Candidate differentially methylated regions (DMRs) associated with 12-month total AOSI score, an indicator of autism risk.

Rank	Chr	Start	End	# Probes	Average MethylationQ1	Avg.Meth.Q4	Q4-Q1 difference	Smoothed effect estimate	FWER*P-value	Gene symbol	Location relative to gene	Distance from transcription start site
1	15	25467684	25471631	51	0.469	0.275	−0.194	−0.0130	0	SNORD115-15	Inside intron	16 276
2	15	25479 139	25482453	44	0.478	0.281	−0.197	−0.0137	0	SNORD115-11	Covers	0
3	15	25448132	25450278	27	0.652	0.426	−0.226	−0.0161	0	SNORD115-17	Covers	0
4	1	246058026	246059550	23	0.473	0.689	0.216	0.0181	0	SMYD3	Inside intron	521160
5	7	100678945	100682385	32	0.643	0.500	−0.143	−0.0126	0	MUC17	Inside exon	15582
6	12	114527770	114529240	21	0.485	0.252	−0.233	−0.0192	0	RBM19	Upstream	123595
7	4	89679863	89681417	23	0.625	0.379	−0.245	−0.0173	0	FAM13A	Covers exon(s)	62984
8	4	22389344	22391462	31	0.559	0.439	−0.119	−0.0124	0	GPR125	Covers exon(s)	126209
9	4	10167294	10169448	32	0.490	0.353	−0.137	−0.0119	0	WDR1	Upstream	48722
10	15	25445204	25447276	28	0.665	0.492	−0.174	−0.0132	1.00E-04	SNORD115-17	Covers	0

FIG. 5

Table 3: Gene ontology, biological processes enriched in differentially methylated regions associated with 12-month AOSI score

Rank	GOBPID	P-value	Odds ratio	Expected Count	Count	Size	Term
1	GO:0006928	0.00027	3.55	5.0	14	1495	Cellular component movement
2	GO:0040011	0.00035	3.61	4.5	13	1344	Locomotion
3	GO:0030155	0.00045	6.86	1.0	6	297	Regulation of cell adhesion
4	GO:0007155	0.00139	3.57	3.3	10	989	Cell adhesion
5	GO:0022610	0.00143	3.56	3.3	10	993	Biological adhesion
6	GO:0051056	0.00161	5.31	1.3	6	380	Regulation of small gtpase mediated signal transduction
7	GO:0022008	0.00226	3.15	4.2	11	1242	Neurogenesis
8	GO:0048011	0.00231	5.96	0.9	5	278	Neurotrophin TRK receptor signaling pathway
9	GO:0038179	0.00242	5.89	0.9	5	281	Neurotrophin signaling pathway
10	GO:0007626	0.00244	7.77	0.6	4	169	Locomotory behavior
11	GO:0008543	0.00325	7.15	0.6	4	183	Fibroblast growth factor receptor signaling pathway
12	GO:0044344	0.00462	6.46	0.7	4	202	Cellular response to fibroblast growth factor stimulus
13	GO:0071774	0.00478	6.39	0.7	4	204	Response to fibroblast growth factor
14	GO:0048699	0.00485	2.96	3.9	10	1173	Generation of neurons
15	GO:0007264	0.00618	3.99	1.7	6	499	Small gtpase mediated signal transduction
16	GO:0007399	0.00647	2.52	6.2	13	1848	Nervous system development
17	GO:0051235	0.00653	5.83	0.7	4	223	Maintenance of location
18	GO:0007173	0.00663	5.80	0.8	4	224	Epidermal growth factor receptor signaling pathway
19	GO:0009894	0.00687	3.47	2.3	7	675	Regulation of catabolic process
20	GO:0007265	0.00695	5.72	0.8	4	227	Ras protein signal transduction
21	GO:0038127	0.00706	5.70	0.8	4	228	ERBB signaling pathway
22	GO:0007268	0.00726	3.44	2.3	7	682	Synaptic transmission
23	GO:0007411	0.00748	4.47	1.2	5	367	Axon guidance

FIG. 6

Rank	GOBPID	P-value	Odds ratio	Expected Count	Count	Size	Term
24	GO:0097485	0.00748	4.47	1.2	5	367	Neuron projection guidance
25	GO:0051674	0.00802	3.37	2.3	7	695	Localization of cell
26	GO:0050790	0.00841	2.50	5.7	12	1688	Regulation of catalytic activity
27	GO:0030182	0.00891	2.84	3.6	9	1081	Neuron differentiation

GOBPID, Gene Ontology Biological Processes ID.

FIG. 6 (Con't.)

Table 4. Association between SVs and variables for CHARM samples

	P.SV1	P.SV2	P.SV3	P.SV4	P.SV5	P.SV6	P.SV7	P.SV8	P.SV9	P.SV10
AOSI	6.40E-02	9.80E-01	6.10E-03	3.30E-01	5.40E-01	6.90E-01	9.60E-01	3.40E-01	5.90E-01	4.80E-01
Age	6.50E-01	9.60E-04	8.80E-01	7.90E-01	7.50E-01	8.60E-01	3.80E-01	5.60E-01	6.30E-01	3.50E-01
Race/Ethnicity	2.60E-01	5.70E-01	2.50E-01	7.50E-01	3.00E-01	3.90E-01	1.90E-01	6.50E-02	5.90E-01	3.60E-01
Study Site	5.00E-01	6.90E-01	9.50E-01	8.50E-01	4.60E-01	5.70E-01	9.50E-01	2.60E-02	2.00E-01	4.80E-01
Paternity Status	6.90E-01	2.60E-01	7.50E-01	9.20E-01	2.70E-01	3.60E-01	5.90E-01	8.30E-01	4.70E-02	6.10E-01
Education	8.10E-02	8.40E-01	9.50E-01	5.40E-01	3.20E-02	3.40E-01	9.30E-01	2.40E-01	4.10E-02	5.00E-01
Income	3.70E-01	2.00E-01	5.10E-01	4.00E-01	5.70E-01	8.20E-01	1.90E-01	8.40E-01	4.90E-03	1.70E-02
Hybridization Date	1.50E-02	5.10E-04	3.40E-01	5.70E-04	6.60E-02	5.00E-01	1.30E-01	5.40E-01	9.80E-01	4.50E-01
Shearing Date	1.80E-02	2.60E-03	2.60E-01	1.80E-01	5.10E-01	2.10E-04	6.50E-02	1.70E-01	4.50E-01	7.10E-01
Hydroshear Machine	5.80E-01	1.20E-02	7.20E-01	3.00E-01	6.20E-01	5.10E-01	3.80E-01	7.60E-01	2.40E-01	5.00E-01
CHARM Gel ID	8.10E-02	4.00E-03	8.60E-01	6.00E-06	1.10E-01	5.10E-04	4.70E-02	8.70E-02	6.10E-02	5.20E-01
Gel Location	8.50E-01	5.10E-01	9.10E-01	9.10E-01	2.60E-01	9.30E-01	3.90E-01	8.80E-01	6.60E-01	5.40E-01

FIG. 7

Table 5. Association between SVs and variables for 450k samples

	P.SV1	P.SV2	P.SV3	P.SV4	P.SV5	P.SV6	P.SV7
AOSI	8.10E-02	2.10E-01	5.60E-01	8.30E-01	6.20E-01	4.40E-01	9.90E-01
Age	4.20E-01	6.90E-01	2.10E-01	8.10E-01	5.30E-01	1.20E-02	2.00E-01
Race/Ethnicity	8.10E-02	4.10E-01	3.10E-01	4.70E-01	1.20E-03	1.10E-01	1.80E-01
Study Site	5.20E-01	7.20E-01	1.20E-01	4.40E-01	6.60E-01	5.20E-01	3.00E-01
Paternity Status	7.70E-01	1.10E-01	5.80E-01	6.70E-01	9.50E-01	1.60E-01	7.20E-01
Education	8.90E-01	4.10E-01	5.40E-01	8.90E-01	1.10E-01	3.60E-02	9.80E-01
Income	3.30E-01	2.50E-01	1.10E-02	7.50E-01	7.10E-01	1.60E-01	6.30E-01
Hybridization Date	2.80E-02	2.60E-04	3.20E-01	4.00E-01	1.80E-01	5.80E-01	6.80E-01

FIG. 8

chr	start	end	pns	*indexStart*	indexEnd	name	annotation	description	region	distance	subregion	Inside distance	meanQ1	meanQ4	Avg. Meth. Q4-Q1
chr15	**25467684**	**25471631**	76056	*1141218*	1141268	SNORD115-15	NR_003307	inside intron	inside	16276	inside intron	5926	0.469151	0.275354	-0.194
chr15	**25479139**	**25482453**	76059	*1141348*	1141391	SNORD115-11	NR_003303	covers	covers	0	NA	NA	0.477932	0.280985	-0.197
chr15	**25448132**	**25450278**	76052	*1141000*	1141026	SNORD115-17	NR_003309	covers	covers	0	NA	NA	0.651876	0.426209	-0.226
chr1	**2.46E+08**	**2.46E+08**	21264	*313791*	313813	SMYD3	NM_022743	inside intron	inside	521160	inside intron	-19281	0.472818	0.688592	0.216
chr7	**1.01E+08**	**1.01E+08**	223232	*3291888*	3291919	MUC17	NM_001040105	inside exon	inside	15582	inside exon	0	0.643069	0.500319	-0.143
chr12	**1.15E+08**	**1.15E+08**	57148	*863488*	863508	RBM19	NM_016196	upstream	upstream	123595	NA	NA	0.485274	0.251892	-0.233
chr4	**89679863**	**89681417**	184620	*2738285*	2738307	FAM13A	NM_001015045	covers exon(s)	inside	62984	covers exon(s)	0	0.624749	0.379259	-0.245
chr4	**22389344**	**22391462**	182022	*2705346*	2705376	GPR125	NM_145290	covers exon(s)	inside	126209	covers exon(s)	0	0.558568	0.439257	-0.119
chr4	**10167294**	**10169448**	181412	*2698042*	2698073	WDR1	NM_005112	upstream	upstream	48722	NA	NA	0.489751	0.352564	-0.137
chr15	**25445204**	**25447276**	76052	*1140962*	1140989	SNORD115-17	NR_003309	covers	covers	0	NA	NA	0.665443	0.491772	-0.174
chr15	**25463442**	**25465002**	76055	*1141165*	1141187	SNORD115-15	NR_003307	inside intron	inside	12034	inside intron	-12012	0.528426	0.294331	-0.234
chr14	**88156779**	**88158342**	72201	*1078432*	1078454	GALC	NM_000153	downstream	downstream	301564	NA	NA	0.54206	0.384889	-0.157
chr8	**1.37E+08**	**1.37E+08**	239105	*3523391*	3523415	KHDRBS3	NM_006558	downstream	downstream	408932	NA	NA	0.745972	0.635565	-0.11
chr8	**1.33E+08**	**1.33E+08**	238632	*3517039*	3517062	HHLA1	NM_001145095	covers exon(s)	inside	27062	covers exon(s)	0	0.693966	0.53793	-0.156
chr8	**1.36E+08**	**1.36E+08**	239045	*3522602*	3522628	LOC286094	NR_026706	downstream	downstream	116955	NA	NA	0.629959	0.526785	-0.103
chr12	**1.15E+08**	**1.15E+08**	57147	*863469*	863487	RBM19	NM_016196	upstream	upstream	112839	NA	NA	0.636481	0.439491	-0.197
chr5	**1.22E+08**	**1.22E+08**	196230	*2900782*	2900793	PRDM6	NM_001136239	inside intron	inside	22607	inside intron	-11792	0.565367	0.295285	-0.27
chr15	**98822899**	**98824463**	84056	*1258905*	1258927	FAM169B	NM_182562	downstream	downstream	233147	NA	NA	0.542908	0.665479	0.123
chr2	**1.25E+08**	**1.25E+08**	137471	*2047549*	2047571	CNTNAP5	NM_130773	inside intron	inside	239138	inside intron	-22032	0.619582	0.47731	-0.142
chr15	**25424890**	**25426900**	76049	*1140737*	1140761	SNORD115-6	NR_003298	covers	covers	0	NA	NA	0.686641	0.528487	-0.158
chr7	**1.56E+08**	**1.56E+08**	227891	*3360170*	3360196	C7orf4	NR_024119	upstream	upstream	341723	NA	NA	0.687903	0.770717	0.083
chr22	**17898315**	**17899836**	158888	*2366859*	2366879	CECR2	NM_031413	upstream	upstream	56791	NA	NA	0.474008	0.349029	-0.125
chrX	**63444648**	**63446044**	255314	*3760336*	3760356	ASB12	NM_130388	overlaps 5'	overlaps 5'	0	overlaps exon downstream	0	0.491441	0.357414	-0.134
chr12	**1.15E+08**	**1.15E+08**	57159	*863628*	863642	TBX5	NM_181486	downstream	downstream	240248	NA	NA	0.654116	0.46388	-0.19
chr4	**74368982**	**74369684**	183922	*2729456*	2729466	AFM	NM_001133	overlaps exon upstream	inside	21521	overlaps exon upstream	0	0.413221	0.616191	0.203

FIG. 9

chr	start	end	pns	*indexStart*	indexEnd	name	annotation	description	region	distance	subregion	Inside distance	meanQ1	meanQ4	Avg. Meth. Q4-Q1
chr6	**4426786**	**4428049**	202983	*2998721*	2998739	CDYL	NR_026590	upstream	upstream	278343	NA	NA	0.684557	0.858445	0.174
chr8	**1.41E+08**	**1.41E+08**	239386	*3527011*	3527029	KCNK9	NM_016601	downstream	downstream	206374	NA	NA	0.761926	0.647451	-0.114
chr6	**1.67E+08**	**1.67E+08**	214163	*3158375*	3158389	RPS6KA2	NM_001006932	inside intron	inside	76502	inside intron	-72419	0.547301	0.767905	0.221
chr15	**25486087**	**25487319**	76060	*1141432*	1141449	SNORD115-39	NR_003354	covers	covers	0	NA	NA	0.49737	0.312202	-0.185
chr15	**25493410**	**25495060**	76063	*1141504*	1141522	SNORD115-11	NR_003303	covers	covers	0	NA	NA	0.515475	0.335426	-0.18
chr12	**1.31E+08**	**1.31E+08**	59774	*899396*	899417	FZD10	NM_007197	downstream	downstream	22817	NA	NA	0.33892	0.500125	0.161
chr20	**57611918**	**57613130**	153335	*2276713*	2276730	SLMO2	NM_016045	covers exon(s)	inside	4770	covers exon(s)	0	0.728206	0.558234	-0.17
chr12	**1.15E+08**	**1.15E+08**	57154	*863563*	863578	RBM19	NM_016196	upstream	upstream	183049	NA	NA	0.75416	0.630376	-0.124
chrX	**1.19E+08**	**1.19E+08**	256820	*3780838*	3780857	NKAP	NM_024528	upstream	upstream	52119	NA	NA	0.662124	0.531413	-0.131
chr17	**50949540**	**50950400**	103642	*1545856*	1545868	CA10	NM_001082534	upstream	upstream	712164	NA	NA	0.339707	0.583866	0.244
chr4	**78343908**	**78344958**	184111	*2731976*	2731991	CXCL13	NM_006419	upstream	upstream	87948	NA	NA	0.807468	0.665705	-0.142
chr11	**1.32E+08**	**1.32E+08**	48528	*736850*	736865	NTM	NM_001048209	inside intron	inside	498551	inside intron	41462	0.620052	0.470138	-0.15
chr8	**1.41E+08**	**1.41E+08**	239387	*3527030*	3527048	KCNK9	NM_016601	downstream	downstream	200743	NA	NA	0.653724	0.534763	-0.119
chr4	**1.85E+08**	**1.85E+08**	188433	*2787782*	2787798	ENPP6	NM_153343	inside intron	inside	10235	inside intron	-9853	0.317927	0.465539	0.148
chr12	**1.15E+08**	**1.15E+08**	57155	*863579*	863594	RBM19	NM_016196	upstream	upstream	187314	NA	NA	0.802292	0.676329	-0.126
chr6	**36328216**	**36329200**	206002	*3043441*	3043455	ETV7	NM_016135	downstream	downstream	26266	NA	NA	0.535103	0.648506	0.113
chr18	**77409595**	**77410720**	115500	*1728156*	1728171	CTDP1	NM_048368	upstream	upstream	29080	NA	NA	0.611699	0.806933	0.195
chr8	**9012834**	**9013562**	230336	*3403593*	3403603	PPP1R3B	NM_024607	upstream	upstream	4615	NA	NA	0.587909	0.820327	0.232
chr8	**3117588**	**3118214**	229525	*3391487*	3391496	CSMD1	NM_033225	inside intron	inside	1734113	inside intron	-23451	0.593093	0.285982	-0.307
chr10	**1.14E+08**	**1.14E+08**	32209	*475914*	475932	GUCY2G	NR_028134	covers exon(s)	inside	41458	covers exon(s)	0	0.77689	0.558009	-0.219
chr5	**7078325**	**7079788**	191030	*2832095*	2832116	ADCY2	NM_020546	upstream	upstream	316554	NA	NA	0.503321	0.371005	-0.132
chr12	**1.15E+08**	**1.15E+08**	57153	*863548*	863562	RBM19	NM_016196	upstream	upstream	171392	NA	NA	0.596119	0.445816	-0.15
chr8	**1.35E+08**	**1.35E+08**	238914	*3520884*	3520901	ZFAT	NM_020863	downstream	downstream	421791	NA	NA	0.576324	0.47978	-0.097
chr15	**29905675**	**29906927**	76679	*1152007*	1152025	FAM189A1	NM_015307	upstream	upstream	42749	NA	NA	0.643865	0.526213	-0.118
chr8	**1.35E+08**	**1.35E+08**	238900	*3520729*	3520751	ST3GAL1	NM_003033	upstream	upstream	342655	NA	NA	0.866104	0.779633	-0.086
chr4	**10161978**	**10163027**	181410	*2698013*	2698028	WDR1	NM_005112	upstream	upstream	43406	NA	NA	0.492097	0.344014	-0.148

FIG. 9 (Con't.)

chr	start	end	pns	indexStart	indexEnd	name	annotation	description	region	distance	subregion	Inside distance	meanQ1	meanQ4	Avg. Meth. Q4-Q1
chr2	**2E+08**	**2E+08**	141521	*2101989*	2102009	FLJ32063	NR_026830	downstream	downstream	113122	NA	NA	0.528912	0.406654	-0.122
chr1	**78605901**	**78606739**	10324	*153883*	153895	GIPC2	NM_017655	downstream	downstream	94313	NA	NA	0.76075	0.56372	-0.197
chr15	**25459846**	**25461182**	76054	*1141132*	1141148	SNORD115-15	NR_003307	inside intron	inside	8438	inside intron	-8416	0.492861	0.309651	-0.183
chr8	**1.41E+08**	**1.41E+08**	239390	*3527075*	3527093	KCNK9	NM_016601	downstream	downstream	186551	NA	NA	0.5303	0.434608	-0.096
chr13	**1.07E+08**	**1.07E+08**	66015	*985627*	985638	EFNB2	NM_004093	downstream	downstream	415445	NA	NA	0.584865	0.422105	-0.163
chr15	**25489676**	**25491034**	76062	*1141468*	1141483	SNORD115-41	NR_003356	covers	covers	0	NA	NA	0.559666	0.380419	-0.179
chr12	**43002099**	**43003456**	52252	*794966*	794985	PRICKLE1	NM_153026	upstream	upstream	18528	NA	NA	0.693658	0.785913	0.092
chrX	**1.02E+08**	**1.02E+08**	256189	*3772555*	3772571	RAB40AL	NM_001031834	downstream	downstream	40907	NA	NA	0.442697	0.563504	0.121
chr21	**41955901**	**41956607**	157109	*2336392*	2336402	DSCAM	NM_001389	inside intron	inside	262431	inside intron	-108128	0.643795	0.418844	-0.225
chr8	**1.36E+08**	**1.36E+08**	239035	*3522489*	3522507	LOC286094	NR_026706	upstream	upstream	64761	NA	NA	0.656638	0.58573	-0.071
chr17	**7689320**	**7690653**	98156	*1463162*	1463179	DNAH2	NM_020877	covers exon(s)	inside	66282	covers exon(s)	0	0.62321	0.756548	0.133
chr11	**4496676**	**4498010**	37302	*567240*	567259	OR52K1	NM_001005171	upstream	upstream	12120	NA	NA	0.77034	0.673137	-0.097
chr13	**51058483**	**51059690**	63497	*953301*	953318	DLEU7	NM_198989	downstream	downstream	358194	NA	NA	0.499406	0.379707	-0.12
chr14	**64650774**	**64651424**	70115	*1049298*	1049307	SYNE2	NM_182914	inside intron	inside	331092	inside intron	1717	0.446211	0.650441	0.204
chr19	**55515148**	**55516198**	125158	*1860912*	1860927	NLRP2	NM_017852	downstream	downstream	37378	NA	NA	0.535234	0.657297	0.122
chr18	**3747209**	**3748396**	108864	*1627460*	1627477	DLGAP1	NM_001003809	inside intron	inside	96899	inside intron	4698	0.806479	0.635869	-0.171
chr4	**10185969**	**10187382**	181414	*2698086*	2698106	WDR1	NM_005112	upstream	upstream	67397	NA	NA	0.324001	0.198952	-0.125
chr16	**19893277**	**19893979**	88247	*1319159*	1319169	GPRC5B	NM_016235	inside intron	inside	2171	inside intron	-2069	0.507611	0.322115	-0.185
chr10	**43731078**	**43732120**	26564	*394614*	394629	RASGEF1A	NM_145313	inside intron	inside	30246	inside intron	29509	0.587257	0.444434	-0.143
chr12	**15110580**	**15111874**	51178	*781857*	781875	ARHGDIB	NM_001175	inside intron	inside	2687	inside intron	-2596	0.643189	0.755652	0.112
chr15	**25472045**	**25473445**	76056	*1141274*	1141292	SNORD115-15	NR_003307	inside intron	inside	20637	inside intron	4112	0.438508	0.292147	-0.146
chr2	**198513**	**199670**	125845	*1869964*	1869980	SH3YL1	NM_001159597	downstream	downstream	64397	NA	NA	0.474226	0.299069	-0.175
chr12	**1.15E+08**	**1.15E+08**	57256	*865214*	865223	TBX3	NM_005996	upstream	upstream	81629	NA	NA	0.495909	0.30914	-0.187
chr4	**10138667**	**10139924**	181403	*2697866*	2697884	WDR1	NM_005112	upstream	upstream	20095	NA	NA	0.426749	0.303735	-0.123
chr15	**25436464**	**25437468**	76050	*1140867*	1140881	SNORD115-5	NR_003297	covers	covers	0	NA	NA	0.646079	0.473793	-0.172
chr2	19994965	19996351	128809	*1919316*	1919335	TTC32	NM_001008237	downstream	downstream	105392	NA	NA	0.82452	0.721415	-0.103

FIG. 9 (Con't.)

chr	start	end	pns	*indexStart*	indexEnd	name	annotation	description	region	distance	subregion	Inside distance	meanQ1	meanQ4	Avg. Meth. Q4-Q1
chr4	**78324835**	**78325656**	184110	*2731963*	2731975	CXCL13	NM_006419	upstream	upstream	107250	NA	NA	0.435902	0.257439	-0.178
chr4	**6327371**	**6328424**	180515	*2682079*	2682093	PPP2R2C	NM_181876	inside intron	inside	55172	inside intron	2052	0.309812	0.512092	0.202
chr4	**10131552**	**10132616**	181400	*2697803*	2697818	WDR1	NM_005112	upstream	upstream	12980	NA	NA	0.720972	0.582453	-0.139
chr12	**1.15E+08**	**1.15E+08**	57157	*863603*	863617	RBM19	NM_016196	upstream	upstream	190776	NA	NA	0.569632	0.43497	-0.135
chr15	**25430262**	**25431120**	76049	*1140798*	1140809	SNORD115-9	NR_003301	covers	covers	0	NA	NA	0.693541	0.483274	-0.21
chr12	**25397742**	**25398595**	51604	*786863*	786875	KRAS	NM_004985	covers exon(s)	inside	5258	covers exon(s)	0	0.807686	0.655459	-0.152
chr12	**1.32E+08**	**1.32E+08**	60095	*904823*	904831	LOC116437	NR_026670	downstream	downstream	228866	NA	NA	0.745722	0.483933	-0.262
chr12	**77965737**	**77966707**	54635	*830924*	830938	NAV3	NM_014903	upstream	upstream	258361	NA	NA	0.727861	0.809166	0.081
chr8	**6827786**	**6829096**	230181	*3401474*	3401492	DEFA10P	NR_029386	promoter	promoter	1152	NA	NA	0.865968	0.780144	-0.086
chr3	**58671756**	**58672720**	170838	*2549185*	2549199	FAM3D	NM_138805	upstream	upstream	19196	NA	NA	0.409253	0.265541	-0.144
chr8	**1.41E+08**	**1.41E+08**	239389	*3527061*	3527074	KCNK9	NM_016601	downstream	downstream	188599	NA	NA	0.694405	0.579442	-0.115
chr12	**1.15E+08**	**1.15E+08**	57152	*863535*	863547	RBM19	NM_016196	upstream	upstream	168802	NA	NA	0.702969	0.578965	-0.124
chr2	**2.37E+08**	**2.37E+08**	144818	*2148946*	2148958	AGAP1	NM_001037131	inside intron	inside	271537	inside intron	-15141	0.739001	0.884471	0.145
chr5	**5509981**	**5511233**	190731	*2827095*	2827113	KIAA0947	NM_015325	downstream	downstream	87175	NA	NA	0.679047	0.55901	-0.12
chr8	**6032892**	**6033931**	230006	*3398674*	3398689	MCPH1	NM_024596	upstream	upstream	230189	NA	NA	0.545149	0.705817	0.161
chr15	**25465300**	**25466304**	76055	*1141190*	1141204	SNORD115-15	NR_003307	inside intron	inside	13892	inside intron	11253	0.485419	0.310513	-0.175
chr9	**25751846**	**25752987**	242377	*3571686*	3571702	TUSC1	NM_001004125	upstream	upstream	72991	NA	NA	0.478375	0.616756	0.138
chr15	**25432438**	**25433678**	76050	*1140822*	1140835	SNORD115-10	NR_003302	covers	covers	0	NA	NA	0.585252	0.412302	-0.173
chr7	**1.56E+08**	**1.56E+08**	227890	*3360142*	3360155	C7orf4	NR_024119	upstream	upstream	345566	NA	NA	0.649331	0.744946	0.096
chr15	**29517550**	**29518244**	76623	*1151152*	1151161	FAM189A1	NM_015307	inside intron	inside	344682	inside intron	-26327	0.364246	0.584918	0.221
chr4	**22265556**	**22266590**	182021	*2705330*	2705345	GPR125	NM_145290	downstream	downstream	251081	NA	NA	0.429441	0.330209	-0.099
chr8	**1.36E+08**	**1.36E+08**	239038	*3522530*	3522542	LOC286094	NR_026706	inside intron	inside	1199	inside intron	-814	0.644605	0.543119	-0.101
chr17	**53397793**	**53398901**	103702	*1546595*	1546609	HLF	NM_002126	overlaps exon upstream	inside	55473	overlaps exon upstream	0	0.74192	0.640123	-0.102
chr20	**52778923**	**52779623**	152544	*2263887*	2263897	CYP24A1	NM_001128915	covers exon(s)	inside	10892	covers exon(s)	0	0.617865	0.424405	-0.193
chr5	**5146252**	**5147358**	190661	*2825729*	2825745	ADAMTS16	NM_139056	overlaps exon downstream	inside	5810	overlaps exon downstream	0	0.581875	0.454515	-0.127
chr5	**8570869**	**8571655**	191191	*2834492*	2834503	SEMA5A	NM_003966	downstream	downstream	974577	NA	NA	0.702856	0.537334	-0.166

FIG. 9 (Con't.)

chr	start	end	pns	*indexStart*	indexEnd	name	annotation	description	region	distance	subregion	Inside distance	meanQ1	meanQ4	Avg. Meth. Q4-Q1
chr15	**26220504**	**26221062**	76167	*1143304*	1143312	ATP10A	NM_024490	upstream	upstream	112156	NA	NA	0.37728	0.558655	0.181
chr8	**1.37E+08**	**1.37E+08**	239099	*3523337*	3523348	KHDRBS3	NM_006558	downstream	downstream	323468	NA	NA	0.7429	0.611102	-0.132
chr15	**25482945**	**25483985**	76059	*1141397*	1141411	SNORD115-37	NR_003352	covers	covers	0	NA	NA	0.509899	0.334602	-0.175
chr7	**84859824**	**84860704**	222210	*3278257*	3278269	SEMA3D	NM_152754	upstream	upstream	108578	NA	NA	0.520053	0.38477	-0.135
chr4	**10163375**	**10164249**	181411	*2698029*	2698041	WDR1	NM_005112	upstream	upstream	44803	NA	NA	0.495863	0.33853	-0.157
chr8	**1.37E+08**	**1.37E+08**	239101	*3523357*	3523367	KHDRBS3	NM_006558	downstream	downstream	374983	NA	NA	0.80075	0.670524	-0.13
chr15	**25442434**	**25443294**	76052	*1140932*	1140944	SNORD115-16	NR_003308	promoter	promoter	1300	NA	NA	0.639393	0.464373	-0.175
chr2	**38974353**	**38975234**	130809	*1949007*	1949019	SFRS7	NM_001031684	overlaps exon downstream	inside	3401	overlaps exon downstream	0	0.667571	0.516251	-0.151
chr5	**1.23E+08**	**1.23E+08**	196269	*2901322*	2901338	CSNK1G3	NM_001031812	downstream	downstream	448680	NA	NA	0.521254	0.612127	0.091
chr11	**43045431**	**43046401**	39766	*601825*	601839	API5	NM_001142931	upstream	upstream	287103	NA	NA	0.583307	0.483387	-0.1
chr7	**1.56E+08**	**1.56E+08**	227889	*3360128*	3360141	C7orf4	NR_024119	upstream	upstream	351320	NA	NA	0.672055	0.757141	0.085
chr9	**19434828**	**19435390**	242123	*3568540*	3568548	ACER2	NM_001010887	covers exon(s)	inside	25904	covers exon(s)	0	0.611054	0.796888	0.186
chr8	**1.36E+08**	**1.36E+08**	239030	*3522440*	3522454	LOC286094	NR_026706	upstream	upstream	102360	NA	NA	0.809549	0.71462	-0.095
chr1	**5169493**	**5170281**	1291	*22383*	22394	AJAP1	NM_018836	downstream	downstream	454389	NA	NA	0.574986	0.406699	-0.168
chr17	**46137266**	**46138284**	103022	*1535556*	1535570	NFE2L1	NM_003204	inside exon	inside	11581	inside exon	0	0.5792	0.466649	-0.113
chr15	**25475757**	**25476977**	76057	*1141316*	1141331	SNORD115-15	NR_003307	inside intron	inside	24349	inside intron	580	0.434875	0.291977	-0.143
chr10	**1.31E+08**	**1.31E+08**	35081	*525785*	525798	MGMT	NM_002412	upstream	upstream	144092	NA	NA	0.478635	0.58273	0.104
chr14	**72740541**	**72741091**	70921	*1060777*	1060785	RGS6	NM_004296	inside intron	inside	340756	inside intron	77711	0.394993	0.19212	-0.203
chr11	**62996820**	**62997502**	41425	*626327*	626337	SLC22A25	NM_199352	overlaps 5'	overlaps 5'	0	inside exon	0	0.542791	0.365753	-0.177
chr2	**1358180**	**1358956**	126137	*1875604*	1875612	SNTG2	NM_018968	inside intron	inside	411627	inside intron	12158	0.585482	0.402988	-0.182
chr20	**61572412**	**61573256**	154374	*2296318*	2296330	C20orf11	NM_017896	covers exon(s)	inside	2972	covers exon(s)	0	0.577654	0.445734	-0.132
chr1	**1.78E+08**	**1.78E+08**	15112	*223484*	223498	SEC16B	NM_033127	upstream	upstream	35130	NA	NA	0.627165	0.516967	-0.11
chr6	**1.57E+08**	**1.57E+08**	212488	*3130174*	3130187	ARID1B	NM_017519	inside intron	inside	389547	inside intron	-314	0.606934	0.477223	-0.13
chr8	**1.41E+08**	**1.41E+08**	239388	*3527049*	3527060	KCNK9	NM_016601	downstream	downstream	189985	NA	NA	0.655081	0.531569	-0.124
chr15	**25484679**	**25485713**	76060	*1141414*	1141427	SNORD115-38	NR_003353	covers	covers	0	NA	NA	0.472444	0.308097	-0.164
chr19	**9028995**	**9029735**	118123	*1761381*	1761391	MUC16	NM_024690	inside intron	inside	62282	inside intron	600	0.69412	0.558414	-0.136

FIG. 9 (Con't.)

chr	start	end	pns	indexStart	indexEnd	name	annotation	description	region	distance	subregion	Inside distance	meanQ1	meanQ4	Avg. Meth. Q4-Q1
chr7	1.01E+08	1.01E+08	223233	3291920	3291931	MUC17	NM_001040105	inside exon	inside	19540	inside exon	0	0.468859	0.326129	-0.143
chr4	32323223	32323693	182573	2712639	2712646	PCDH7	NM_032457	downstream	downstream	1601187	NA	NA	0.586063	0.380938	-0.205
chr22	27164750	27165574	160473	2391378	2391390	MIAT	NR_003491	downstream	downstream	111267	NA	NA	0.402491	0.529659	0.127
chr8	1.4E+08	1.4E+08	239383	3526966	3526978	KCNK9	NM_016601	downstream	downstream	227658	NA	NA	0.738494	0.640918	-0.098
chr21	41428841	41429669	157036	2335296	2335308	DSCAM	NM_001389	inside intron	inside	789369	inside intron	1075	0.450637	0.60486	0.154
chr4	1.9E+08	1.9E+08	189296	2800042	2800056	FRG1	NM_004477	upstream	upstream	465345	NA	NA	0.506573	0.408481	-0.098
chr2	75246952	75247704	133630	1989859	1989870	TACR1	NM_001058	downstream	downstream	178940	NA	NA	0.745042	0.599051	-0.146
chr7	1.57E+08	1.57E+08	228179	3364936	3364949	DNAJB6	NM_058246	downstream	downstream	129010	NA	NA	0.561707	0.678325	0.117
chr15	25456582	25457452	76053	1141094	1141105	SNORD115-23	NR_003315	covers	covers	0	NA	NA	0.473955	0.287512	-0.186
chr8	1.36E+08	1.36E+08	239039	3522543	3522554	LOC286094	NR_026706	inside intron	inside	34780	inside intron	21508	0.575658	0.471126	-0.105
chr3	86978543	86979173	172555	2571500	2571509	VGLL3	NM_016206	downstream	downstream	61083	NA	NA	0.583431	0.723365	0.14
chr5	7785778	7786404	191120	2833321	2833330	ADCY2	NM_020546	inside intron	inside	389436	inside intron	-1217	0.331376	0.54856	0.217
chr5	72562315	72563475	194223	2874762	2874777	TMEM174	NM_153217	downstream	downstream	93293	NA	NA	0.651953	0.724475	0.073
chrX	1.03E+08	1.03E+08	256267	3773591	3773606	H2BFM	NM_001164416	downstream	downstream	7827	NA	NA	0.7356	0.648812	-0.087
chr4	79428290	79428774	184163	2732637	2732644	FRAS1	NM_025074	covers exon(s)	inside	449567	covers exon(s)	0	0.446489	0.65614	0.21
chr8	1.36E+08	1.36E+08	239034	3522476	3522488	LOC286094	NR_026706	upstream	upstream	66635	NA	NA	0.698084	0.62203	-0.076
chr8	1.4E+08	1.4E+08	239384	3526979	3526987	KCNK9	NM_016601	downstream	downstream	226716	NA	NA	0.658191	0.501106	-0.157
chr11	4439562	4440332	37299	567201	567212	TRIM21	NM_003141	upstream	upstream	24637	NA	NA	0.583138	0.44412	-0.139
chr17	14939913	14940685	99187	1478404	1478415	PMP22	NM_153322	downstream	downstream	223407	NA	NA	0.355868	0.50365	0.148
chr15	25477475	25478313	76058	1141332	1141343	SNORD115-34	NR_003349	covers	covers	0	NA	NA	0.391883	0.228408	-0.163
chr4	66394009	66394849	183753	2727529	2727541	EPHA5	NM_004439	inside intron	inside	140803	inside intron	32749	0.255853	0.374543	0.119
chr3	58449165	58449986	170791	2548582	2548594	KCTD6	NM_001128214	upstream	upstream	27836	NA	NA	0.449515	0.352096	-0.097
chr4	98344854	98345644	184820	2740700	2740711	C4orf37	NM_174952	downstream	downstream	718746	NA	NA	0.364737	0.520265	0.156
chr10	1.12E+08	1.12E+08	32009	473104	473118	MXI1	NM_001008541	inside intron	inside	11412	inside intron	5508	0.423245	0.532468	0.109
chr13	1.1E+08	1.1E+08	66448	991989	992000	IRS2	NM_003749	downstream	downstream	57911	NA	NA	0.442849	0.590573	0.148
chr2	46195882	46196793	131482	1958539	1958552	PRKCE	NM_005400	inside intron	inside	316840	inside intron	6774	0.360913	0.423035	0.062

FIG. 9 (Con't.)

chr	start	end	pns	*indexStart*	indexEnd	name	annotation	description	region	distance	subregion	Inside distance	meanQ1	meanQ4	Avg. Meth. Q4-Q1
chr2	**4702568**	**4703320**	126920	*1890394*	1890405	ALLC	NM_018436	downstream	downstream	996783	NA	NA	0.600055	0.468152	-0.132
chr1	**2.48E+08**	**2.48E+08**	21464	*316611*	316623	C1orf150	NM_145278	promoter	promoter	969	NA	NA	0.567442	0.496792	-0.071
chr8	**41484850**	**41485602**	233874	*3455826*	3455837	AGPAT6	NM_178819	downstream	downstream	49144	NA	NA	0.713018	0.834638	0.122
chr12	**1.15E+08**	**1.15E+08**	57158	*863618*	863627	TBX5	NM_181486	downstream	downstream	241638	NA	NA	0.568007	0.417364	-0.151
chr9	**89251670**	**89252453**	244445	*3600009*	3600020	ZCCHC6	NM_024617	upstream	upstream	282302	NA	NA	0.708059	0.542492	-0.166
chr2	**42565189**	**42565821**	131037	*1951914*	1951922	EML4	NM_001145076	downstream	downstream	168700	NA	NA	0.70613	0.557703	-0.148
chr4	**10133742**	**10134440**	181401	*2697828*	2697838	WDR1	NM_005112	upstream	upstream	15170	NA	NA	0.534871	0.3829	-0.152
chr10	**1.07E+08**	**1.07E+08**	31795	*470676*	470684	SORCS3	NM_014978	inside intron	inside	166488	inside intron	34653	0.8278	0.62612	-0.202
chr9	**97888805**	**97889445**	245659	*3619958*	3619967	FANCC	NM_000136	covers exon(s)	inside	190545	covers exon(s)	0	0.443974	0.298425	-0.146
chr2	**38975737**	**38976370**	130810	*1949020*	1949029	SFRS7	NM_001031684	overlaps exon upstream	inside	2265	overlaps exon upstream	0	0.5126	0.325801	-0.187
chr8	**58128966**	**58129512**	234946	*3470516*	3470524	C8orf71	NR_026772	upstream	upstream	62589	NA	NA	0.434242	0.609905	0.176
chr8	**1.35E+08**	**1.35E+08**	238921	*3520958*	3520972	ZFAT	NM_020863	downstream	downstream	341950	NA	NA	0.611145	0.532514	-0.079
chr11	**1.27E+08**	**1.27E+08**	47813	*725243*	725252	KIRREL3	NM_001161707	upstream	upstream	70615	NA	NA	0.738578	0.53662	-0.202
chrY	**8455694**	**8456464**	258757	*3807528*	3807539	TTTY18	NR_001550	downstream	downstream	95454	NA	NA	0.775131	0.673751	-0.101
chr15	**25458880**	**25459510**	76054	*1141119*	1141128	SNORD115-15	NR_003307	inside intron	inside	7472	inside intron	-7450	0.489149	0.304934	-0.184
chr16	**66684329**	**66685241**	91910	*1367979*	1367992	CMTM4	NM_178818	inside intron	inside	45368	inside intron	13846	0.746892	0.629192	-0.118
chr1	**5355887**	**5356721**	1330	*23022*	23034	AJAP1	NM_018836	downstream	downstream	640783	NA	NA	0.611467	0.48232	-0.129
chr22	**37386774**	**37387704**	161997	*2412923*	2412936	C22orf33	NM_178552	overlaps 3'	overlaps 3'	16172	covers exon(s)	0	0.306945	0.175341	-0.132
chr4	**1.82E+08**	**1.82E+08**	187895	*2779535*	2779546	MGC45800	NR_027107	downstream	downstream	1130861	NA	NA	0.713826	0.873111	0.159
chr1	**2.46E+08**	**2.46E+08**	21152	*311591*	311603	KIF26B	NM_018012	inside intron	inside	262680	inside intron	995	0.645475	0.546042	-0.099
chr5	**1.43E+08**	**1.43E+08**	198109	*2928762*	2928775	HMHB1	NM_021182	upstream	upstream	181402	NA	NA	0.740634	0.636778	-0.104
chr8	**9106053**	**9106601**	230352	*3403788*	3403796	PPP1R3B	NM_024607	upstream	upstream	97834	NA	NA	0.317824	0.46782	0.15
chr12	**5309904**	**5310444**	50183	*766462*	766470	KCNA5	NM_002234	downstream	downstream	156820	NA	NA	0.654884	0.510396	-0.144
chr12	**1.15E+08**	**1.15E+08**	57160	*863643*	863652	TBX5	NM_181486	downstream	downstream	238996	NA	NA	0.783197	0.656781	-0.126
chr4	**78161987**	**78162827**	184106	*2731915*	2731927	CCNG2	NM_004354	downstream	downstream	83631	NA	NA	0.396646	0.324085	-0.073
chr8	**1.36E+08**	**1.36E+08**	239028	*3522411*	3522422	LOC286094	NR_026706	upstream	upstream	110819	NA	NA	0.805789	0.715245	-0.091

FIG. 9 (Con't.)

chr	start	end	pns	*indexStart*	indexEnd	name	annotation	description	region	distance	subregion	Inside distance	meanQ1	meanQ4	Avg. Meth. Q4-Q1
chr21	**40295474**	**40296016**	156855	*2332251*	2332259	ETS2	NM_005239	downstream	downstream	117626	NA	NA	0.433415	0.56442	0.131
chr8	**32617491**	**32618525**	233158	*3445802*	3445817	NRG1	NM_001159996	covers exon(s)	inside	38141	covers exon(s)	0	0.447239	0.335053	-0.112
chr2	**2E+08**	**2E+08**	141492	*2101552*	2101564	SATB2	NM_015265	inside exon	inside	187917	inside exon	0	0.578241	0.443563	-0.135
chr16	**46604938**	**46605796**	89779	*1338645*	1338657	ANKRD26P1	NR_026556	promoter	promoter	1930	NA	NA	0.310104	0.405328	0.095
chr4	**1.9E+08**	**1.9E+08**	189185	*2798543*	2798551	TRIML1	NM_178556	downstream	downstream	472804	NA	NA	0.451397	0.581673	0.13
chr19	**55529476**	**55530094**	125161	*1860958*	1860967	GP6	NM_001083899	covers exon(s)	inside	19537	covers exon(s)	0	0.667752	0.514439	-0.153
chr5	**1.22E+08**	**1.22E+08**	196185	*2900224*	2900236	SNCAIP	NM_005460	covers exon(s)	inside	91019	covers exon(s)	0	0.680312	0.584282	-0.096
chr3	**1.3E+08**	**1.3E+08**	174906	*2604052*	2604058	COL29A1	NM_153264	covers exon(s)	inside	60206	covers exon(s)	0	0.413824	0.654005	0.24
chrX	**1.12E+08**	**1.12E+08**	256568	*3777594*	3777601	ZCCHC16	NM_001004308	inside intron	inside	236905	inside intron	134096	0.628873	0.811628	0.183
chr3	**33686111**	**33686821**	168181	*2508258*	2508268	CLASP2	NM_015097	covers exon(s)	inside	73026	covers exon(s)	0	0.57575	0.439034	-0.137
chr1	**1.78E+08**	**1.78E+08**	15088	*223171*	223182	SEC16B	NM_033127	downstream	downstream	130154	NA	NA	0.777682	0.64649	-0.131
chr22	**35167276**	**35168202**	161611	*2407386*	2407399	ISX	NM_001008494	upstream	upstream	293927	NA	NA	0.604203	0.493148	-0.111

FIG. 9 (Con't.)

chr	start	end	pns	indexStart	indexEnd	name	annotation	description	region	distance	subregion	Inside distance
chr15	25467684	25471631	76056	1141218	1141268	SNORD115-15	NR_003307	inside intron	inside	16276	inside intron	5926
chr15	25479139	25482453	76059	1141348	1141391	SNORD115-11	NR_003303	covers	covers	0	NA	NA
chr15	25448132	25450278	76052	1141000	1141026	SNORD115-17	NR_003309	covers	covers	0	NA	NA
chr1	2.46E+08	2.46E+08	21264	313791	313813	SMYD3	NM_022743	inside intron	inside	521160	inside intron	-19281
chr4	89679863	89681417	184620	2738285	2738307	FAM13A	NM_001015045	covers exon(s)	inside	62984	covers exon(s)	0
chr4	22389344	22391462	182022	2705346	2705376	GPR125	NM_145290	covers exon(s)	inside	126209	covers exon(s)	0
chr15	25445204	25447276	76052	1140962	1140989	SNORD115-17	NR_003309	covers	covers	0	NA	NA
chr15	25463442	25465002	76055	1141165	1141187	SNORD115-15	NR_003307	inside intron	inside	12034	inside intron	-12012
chr15	25424890	25426900	76049	1140737	1140761	SNORD115-6	NR_003298	covers	covers	0	NA	NA
chr7	1.56E+08	1.56E+08	227891	3360170	3360196	C7orf4	NR_024119	upstream	upstream	341723	NA	NA
chr22	17898315	17899836	158888	2366859	2366879	CECR2	NM_031413	upstream	upstream	56791	NA	NA
chrX	63444648	63446044	255314	3760336	3760356	ASB12	NM_130388	overlaps 5'	overlaps 5'	0	overlaps exon downstream	0
chr4	74368982	74369684	183922	2729456	2729466	AFM	NM_001133	overlaps exon upstream	inside	21521	overlaps exon upstream	0
chr6	4426786	4428049	202983	2998721	2998739	CDYL	NR_026590	upstream	upstream	278343	NA	NA
chr6	1.67E+08	1.67E+08	214163	3158375	3158389	RPS6KA2	NM_001006932	inside intron	inside	76502	inside intron	-72419
chr15	25486087	25487319	76060	1141432	1141449	SNORD115-39	NR_003354	covers	covers	0	NA	NA
chr15	25493410	25495060	76063	1141504	1141522	SNORD115-11	NR_003303	covers	covers	0	NA	NA

FIG. 10

chr	start	end	pns	indexStart	indexEnd	name	annotation	description	region	distance	subregion	Inside distance
chr12	1.31E+08	1.31E+08	59774	899396	899417	FZD10	NM_007197	downstream	downstream	22817	NA	NA
chrX	1.19E+08	1.19E+08	256820	3780838	3780857	NKAP	NM_024528	upstream	upstream	52119	NA	NA
chr4	1.85E+08	1.85E+08	188433	2787782	2787798	ENPP6	NM_153343	inside intron	inside	10235	inside intron	-9853
chr18	77409595	77410720	115500	1728156	1728171	CTDP1	NM_048368	upstream	upstream	29080	NA	NA
chr8	9012834	9013562	230336	3403593	3403603	PPP1R3B	NM_024607	upstream	upstream	4615	NA	NA
chr10	1.14E+08	1.14E+08	32209	475914	475932	GUCY2G	NR_028134	covers exon(s)	inside	41458	covers exon(s)	0
chr15	29905675	29906927	76679	1152007	1152025	FAM189A1	NM_015307	upstream	upstream	42749	NA	NA
chr2	2E+08	2E+08	141521	2101989	2102009	FLJ32063	NR_026830	downstream	downstream	113122	NA	NA
chr15	25459846	25461182	76054	1141132	1141148	SNORD115-15	NR_003307	inside intron	inside	8438	inside intron	-8416
chr15	25489676	25491034	76062	1141468	1141483	SNORD115-41	NR_003356	covers	covers	0	NA	NA
chr8	1.36E+08	1.36E+08	239035	3522489	3522507	LOC286094	NR_026706	upstream	upstream	64761	NA	NA
chr17	7689320	7690653	98156	1463162	1463179	DNAH2	NM_020877	covers exon(s)	inside	66282	covers exon(s)	0
chr16	19893277	19893979	88247	1319159	1319169	GPRC5B	NM_016235	inside intron	inside	2171	inside intron	-2069
chr15	25472045	25473445	76056	1141274	1141292	SNORD115-15	NR_003307	inside intron	inside	20637	inside intron	4112
chr2	198513	199670	125845	1869964	1869980	SH3YL1	NM_001159597	downstream	downstream	64397	NA	NA
chr15	25436464	25437468	76050	1140867	1140881	SNORD115-5	NR_003297	covers	covers	0	NA	NA
chr4	6327371	6328424	180515	2682079	2682093	PPP2R2C	NM_181876	inside intron	inside	55172	inside intron	2052
chr15	25430262	25431120	76049	1140798	1140809	SNORD115-9	NR_003301	covers	covers	0	NA	NA
chr8	6827786	6829096	230181	3401474	3401492	DEFA10P	NR_029386	promoter	promoter	1152	NA	NA
chr2	2.37E+08	2.37E+08	144818	2148946	2148958	AGAP1	NM_001037131	inside intron	inside	271537	inside intron	-15141
chr15	25465300	25466304	76055	1141190	1141204	SNORD115-15	NR_003307	inside intron	inside	13892	inside intron	11253
chr15	25432438	25433678	76050	1140822	1140835	SNORD115-10	NR_003302	covers	covers	0	NA	NA
chr5	5146252	5147358	190661	2825729	2825745	ADAMTS16	NM_139056	overlaps	inside	5810	overlaps	0

FIG. 10 (Con't.)

chr	start	end	pns	indexStart	indexEnd	name	annotation	description	region	distance	subregion	Inside distance
								exon downstream			exon downstream	
chr15	25482945	25483985	76059	1141397	1141411	SNORD115-37	NR_003352	covers	covers	0	NA	NA
chr15	25442434	25443294	76052	1140932	1140944	SNORD115-16	NR_003308	promoter	promoter	1300	NA	NA
chr17	46137266	46138284	103022	1535556	1535570	NFE2L1	NM_003204	inside exon	inside	11581	inside exon	0
chr15	25475757	25476977	76057	1141316	1141331	SNORD115-15	NR_003307	inside intron	inside	24349	inside intron	580
chr10	1.31E+08	1.31E+08	35081	525785	525798	MGMT	NM_002412	upstream	upstream	144092	NA	NA
chr11	62996820	62997502	41425	626327	626337	SLC22A25	NM_199352	overlaps 5'	overlaps 5'	0	inside exon	0
chr2	1358180	1358956	126137	1875604	1875612	SNTG2	NM_018968	inside intron	inside	411627	inside intron	12158
chr20	61572412	61573256	154374	2296318	2296330	C20orf11	NM_017896	covers exon(s)	inside	2972	covers exon(s)	0
chr15	25484679	25485713	76060	1141414	1141427	SNORD115-38	NR_003353	covers	covers	0	NA	NA
chr8	1.4E+08	1.4E+08	239383	3526966	3526978	KCNK9	NM_016601	downstream	downstream	227658	NA	NA
chr2	75246952	75247704	133630	1989859	1989870	TACR1	NM_001058	downstream	downstream	178940	NA	NA
chr7	1.57E+08	1.57E+08	228179	3364936	3364949	DNAJB6	NM_058246	downstream	downstream	129010	NA	NA
chr15	25456582	25457452	76053	1141094	1141105	SNORD115-23	NR_003315	covers	covers	0	NA	NA
chr3	86978543	86979173	172555	2571500	2571509	VGLL3	NM_016206	downstream	downstream	61083	NA	NA
chr5	72562315	72563475	194223	2874762	2874777	TMEM174	NM_153217	downstream	downstream	93293	NA	NA
chr8	1.4E+08	1.4E+08	239384	3526979	3526987	KCNK9	NM_016601	downstream	downstream	226716	NA	NA
chr15	25477475	25478313	76058	1141332	1141343	SNORD115-34	NR_003349	covers	covers	0	NA	NA
chr3	58449165	58449986	170791	2548582	2548594	KCTD6	NM_001128214	upstream	upstream	27836	NA	NA
chr10	1.12E+08	1.12E+08	32009	473104	473118	MXI1	NM_001008541	inside intron	inside	11412	inside intron	5508
chr13	1.1E+08	1.1E+08	66448	991989	992000	IRS2	NM_003749	downstream	downstream	57911	NA	NA
chr1	2.48E+08	2.48E+08	21464	316611	316623	C1orf150	NM_145278	promoter	promoter	969	NA	NA

FIG. 10 (Con't.)

chr	start	end	pns	indexStart	indexEnd	name	annotation	description	region	distance	subregion	Inside distance
chr2	42565189	42565821	131037	1951914	1951922	EML4	NM_001145076	downstream	downstream	168700	NA	NA
chr2	38975737	38976370	130810	1949020	1949029	SFRS7	NM_001031684	overlaps exon upstream	inside	2265	overlaps exon upstream	0
chr8	58128966	58129512	234946	3470516	3470524	C8orf71	NR_026772	upstream	upstream	62589	NA	NA
chr15	25458880	25459510	76054	1141119	1141128	SNORD115-15	NR_003307	inside intron	inside	7472	inside intron	-7450
chr16	66684329	66685241	91910	1367979	1367992	CMTM4	NM_178818	inside intron	inside	45368	inside intron	13846
chr1	5355887	5356721	1330	23022	23034	AJAP1	NM_018836	downstream	downstream	640783	NA	NA
chr22	37386774	37387704	161997	2412923	2412936	C22orf33	NM_178552	overlaps 3'	overlaps 3'	16172	covers exon(s)	0
chr1	2.46E+08	2.46E+08	21152	311591	311603	KIF26B	NM_018012	inside intron	inside	262680	inside intron	995
chr5	1.43E+08	1.43E+08	198109	2928762	2928775	HMHB1	NM_021182	upstream	upstream	181402	NA	NA
chr8	9106053	9106601	230352	3403788	3403796	PPP1R3B	NM_024607	upstream	upstream	97834	NA	NA
chr12	5309904	5310444	50183	766462	766470	KCNA5	NM_002234	downstream	downstream	156820	NA	NA
chr8	32617491	32618525	233158	3445802	3445817	NRG1	NM_001159996	covers exon(s)	inside	38141	covers exon(s)	0
chr2	2E+08	2E+08	141492	2101552	2101564	SATB2	NM_015265	inside exon	inside	187917	inside exon	0
chr5	1.22E+08	1.22E+08	196185	2900224	2900236	SNCAIP	NM_005460	covers exon(s)	inside	91019	covers exon(s)	0

FIG. 10 (Con't.)

chr	start	end	pns	indexStart	indexEnd	name	annotation	description	region	distance	subregion	Inside distance	Rank
chr15	25467684	25471631	76056	1141218	1141268	SNORD115-15	NR_003307	inside intron	inside	16276	inside intron	5926	1
chr15	25479139	25482453	76059	1141348	1141391	SNORD115-11	NR_003303	covers	covers	0	NA	NA	2
chr15	25448132	25450278	76052	1141000	1141026	SNORD115-17	NR_003309	covers	covers	0	NA	NA	3
chr15	25445204	25447276	76052	1140962	1140989	SNORD115-17	NR_003309	covers	covers	0	NA	NA	10
chr15	25463442	25465002	76055	1141165	1141187	SNORD115-15	NR_003307	inside intron	inside	12034	inside intron	-12012	11
chr15	25424890	25426900	76049	1140737	1140761	SNORD115-6	NR_003298	covers	covers	0	NA	NA	20
chr22	17898315	17899836	158888	2366859	2366879	CECR2	NM_031413	upstream	upstream	56791	NA	NA	22
chr15	25493410	25495060	76063	1141504	1141522	SNORD115-11	NR_003303	covers	covers	0	NA	NA	30
chr15	25489676	25491034	76062	1141468	1141483	SNORD115-41	NR_003356	covers	covers	0	NA	NA	57
chr15	25436464	25437468	76050	1140867	1140881	SNORD115-5	NR_003297	covers	covers	0	NA	NA	76
chr15	25430262	25431120	76049	1140798	1140809	SNORD115-9	NR_003301	covers	covers	0	NA	NA	82
chr15	25465300	25466304	76055	1141190	1141204	SNORD115-15	NR_003307	inside intron	inside	13892	inside intron	11253	93
chr15	25432438	25433678	76050	1140822	1140835	SNORD115-10	NR_003302	covers	covers	0	NA	NA	95
chr15	25482945	25483985	76059	1141397	1141411	SNORD115-37	NR_003352	covers	covers	0	NA	NA	106
chr11	62996820	62997502	41425	626327	626337	SLC22A25	NM_199352	overlaps 5'	overlaps 5'	0	inside exon	0	122
chr2	75246952	75247704	133630	1989859	1989870	TACR1	NM_001058	downstream	downstream	178940	NA	NA	136
chr15	25456582	25457452	76053	1141094	1141105	SNORD115-23	NR_003315	covers	covers	0	NA	NA	138
chr15	25477475	25478313	76058	1141332	1141343	SNORD115-34	NR_003349	covers	covers	0	NA	NA	149

FIG. 11

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2016/024314

A. CLASSIFICATION OF SUBJECT MATTER
IPC(8) - C12Q 1/68; C40B 30/04; G01N 33/48; G01N 33/50; G01N 33/53; G01N 33/68 (2016.01)
CPC - C12Q 1/6883; C12Q 2600/112; C12Q 2600/118; C12Q 2600/154; G01N 2800/28 (2016.05)
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)
IPC - C12Q 1/68; C40B 30/04; G01N 33/48; G01N 33/50; G01N 33/53; G01N 33/68
CPC - C12Q 1/6883; C12Q 2600/112; C12Q 2600/118; C12Q 2600/154; G01N 2800/28

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
USPC - 435/325; 435/6.16; 435/6.12; 435/6.11; 435/7.9; 435/7.1; 506/9; 506/16 (keyword delimited)

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
PatBase, Google Patents, Google, PubMed
Search terms used: (autism) OR (PDD NOS) OR (pervasive developmental disorder not otherwise specified) OR (Asperger%) (differentially methylated region%) OR DMRs "15q11.2" (SNORD115 "15")

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2010/0125042 A1 (GESCHWIND et al) 20 May 2010 (20.05.2010) entire document	39-41
P, X	FEINBERG et al. "Paternal sperm DNA methylation associated with early signs of autism risk in an autism-enriched cohort," Int. J. Epidemiol. 14 April 2015 (14.04.2015), Vol. 44, Pgs. 1199-1210. entire document	1-12, 14-19, 23, 27, 32, 33, 39-42
A	US 2013/0109583 A1 (BEIM) 02 May 2013 (02.05.2013) entire document	1-12, 14-19, 23, 27, 32, 33, 39-42
A	US 2014/0274748 A1 (MAYO FOUNDATION FOR MEDICAL EDUCATION AND RESEARCH et al) 18 September 2014 (18.09.2014) entire document	1-12, 14-19, 23, 27, 32, 33, 39-42
A	WO 2013/056022 A1 (ZYMO RESEARCH CORPORATION) 18 April 2013 (18.04.2013) entire document	1-12, 14-19, 23, 27, 32, 33, 39-42
A	WO 2007/044780 A2 (EMORY UNIVERSITY et al) 19 April 2007 (19.04.2007) entire document	1-12, 14-19, 23, 27, 32, 33, 39-42
A	WONG et al. "Methylomic analysis of monozygotic twins discordant for autism spectrum disorder and related behavioural traits," Molecular Psychiatry, 23 April 2013 (23.04.2013), Vol. 19, Pgs. 495-503. entire document	1-12, 14-19, 23, 27, 32, 33, 39-42

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* 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	Date of mailing of the international search report
26 July 2016	26 AUG 2016
Name and mailing address of the ISA/ Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, VA 22313-1450 Facsimile No. 571-273-8300	Authorized officer Blaine R. Copenheaver PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2016/024314

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:

see Extra Sheet(s).

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.:
1-12, 14-19, 23, 27, 32, 33, and 39-42 to the extent that they read on SNORD115-15.

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.

Continued from Box No. III Observations where unity of invention is lacking

This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be examined, the appropriate additional examination fees need to be paid.

Group I+: claims 1-42 are drawn differentially methylated regions (DMRs); nucleic acid sequences, a method, and a kit comprising the same.

The first invention of Group I+ is restricted to a differentially methylated region (DMR), a nucleic acid sequence, a method, and a kit comprising the same; wherein then DMR is selected to be SNORD115-15. It is believed that claims 1-12, 14-19, 23, 27, 32, 33, and 39-42 read on this first named invention and thus these claims will be searched without fee to the extent that they read on SNORD115-15.

Applicant is invited to elect additional DMRs for each nucleic acid sequence, method, and/or kit to be searched in a specific combination by paying additional fee for each set of election. An exemplary election would be a differentially methylated region (DMR), a nucleic acid sequence, a method, and a kit comprising the same; wherein then DMR is selected to be SNORD115-11. Additional DMRs will be searched upon the payment of additional fees. Applicants must specify the claims that read on any additional elected inventions. Applicants must further indicate, if applicable, the claims which read on the first named invention if different than what was indicated above for this group. Failure to clearly identify how any paid additional invention fees are to be applied to the "+" group(s) will result in only the first claimed invention to be searched/examined.

The inventions listed in Groups I+ 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 Groups I+ formulas do not share a significant structural element responsible for determining risk of autism spectrum disorder (ASD) in an offspring subject, requiring the selection of alternatives for the differentially methylated region, where "the DMRs are selected from the DMRs as set forth in Table 2, Table, 6, Table 7, or Table 8".

The Groups I+ share the technical features of a method for determining risk of autism spectrum disorder (ASD) in an offspring subject comprising: a) analyzing DNA methylation status in a sample containing sperm from the prospective paternal parent; and b) determining a deviation in the methylation status of (a) as compared to a corresponding normal sample not associated with ASD, wherein a methylation pattern that deviates from the pattern found in the corresponding sample not associated with ASD, is indicative of a risk of ASD in the offspring; a method for determining whether a subject has or is at risk of having autism spectrum disorder (ASD) comprising: a) analyzing DNA methylation status in a DNA sample of the subject; and b) determining a deviation in the methylation status of (a) as compared to a corresponding normal sample not associated with ASD, wherein a methylation pattern that is different from the pattern found in the corresponding sample not associated with ASD, is indicative of a subject having or being at risk of having ASD; a plurality of nucleic acid sequences that selectively hybridizing to a nucleic acid sequence selected from the group consisting of differentially methylated regions (DMRs); a kit for determining whether a subject has or is at risk of having or inheriting a risk of having ASD, comprising: a) a reagent for determining methylation status of a nucleic acid sequence, wherein the nucleic acid sequence is a differentially methylated region (DMR) associated with ASD; and b) instructions for use of the reagent. However, these shared technical features do not represent a contribution over the prior art.

US 2013/0109583 A1 to Beim discloses a method for determining risk of autism spectrum disorder (ASD) in an offspring subject (The invention generally relates to methods and devices for assessing risk to a putative offspring of developing a condition, Abstract; a mutation in the CBS gene affects infertility and is also associated with development of an autism spectrum disorder in the putative offspring., Para. [0008]) comprising: a) analyzing DNA methylation status in a sample containing sperm from the prospective paternal parent; and b) determining a deviation in the methylation status of (a) as compared to a corresponding normal sample not associated with ASD (other embodiments, the biomarker may be an epigenetic factor, such as methylation patterns, Para. [0617], Sperm, Para. [0004]; obtaining a sample from the mammal that includes the infertility- associated biomarker, Para. [0011]; method that can be used to compare mRNA levels in different sample populations to characterize patterns of gene expression, Para. [0670]).

US 2014/0274748 A1 to Mayo Foundation for Medical Education and Research et al. discloses a methylation pattern that deviates from the pattern found in the corresponding sample not associated with disorder (analyzing changes in methylation patterns involves a PCR based process that involves digestion of genomic DNA with methylation-sensitive restriction enzymes, Para. [0015]), is indicative of a risk of the disorder in the offspring; a method for determining whether a subject has or is at risk of having a disorder comprising: a) analyzing DNA methylation status in a DNA sample of the subject (identifying the subject as having a neoplasm when the methylation state of the marker is different than a methylation state of the marker assayed in a subject that does not have a neoplasm, wherein the marker comprises a base in a differentially methylated region (DMR), Para [0031]), and b) determining a deviation in the methylation status of (a) as compared to a corresponding normal sample not associated with a disorder (the methylation state of the marker comprises an increased methylation of the marker relative to a normal methylation state of the marker, Para. [0032]; comparing the methylation state of the marker from the subject sample to a methylation state of the marker from a normal control sample, Para. [0040]; Barrett's esophagus (BE), Para. [0253]); wherein a methylation pattern that is different from the pattern found in the corresponding sample not associated with a disorder, is indicative of a subject having or being at risk of having a disorder (Thus, when methylation in a first population or pool of nucleic acid molecules is different from methylation in a second population or pool of nucleic acid molecules, the methylation state frequency of the first population or pool will be different from the methylation state frequency of the second population or pool., Para. [0081]); a plurality of nucleic acid sequences that selectively hybridizing to a nucleic acid sequence selected from the group consisting of differentially methylated regions (DMRs) (determining the methylation states of multiple markers (such as multiple DMR, Para. [0021]); a kit for determining whether a subject has or is at risk of having or inheriting a risk of having a disorder, comprising: a) a reagent for determining methylation status of a nucleic acid sequence (kit embodiments are provided, e.g., a kit comprising a bisulfite reagent; and a control nucleic acid comprising a sequence from a DMR, Para. [0038]), wherein the nucleic acid sequence is a differentially methylated region (DMR) associated with a disorder; and b) instructions for use of the reagent (in the appropriate containers) and/or supporting materials (e.g., buffers, written instructions, Para. [0114]).

The inventions listed in Groups I+ therefore lack unity under Rule 13 because they do not share a same or corresponding special technical features.

patsnap

专利名称(译)	识别自闭症风险的方法		
公开(公告)号	EP3274477A4	公开(公告)日	2018-08-15
申请号	EP2016773861	申请日	2016-03-25
[标]申请(专利权)人(译)	约翰霍普金斯大学		
申请(专利权)人(译)	约翰·霍普金斯大学		
当前申请(专利权)人(译)	约翰·霍普金斯大学		
[标]发明人	FEINBERG ANDREW P FALLIN DANIELE M		
发明人	FEINBERG, ANDREW, P. FALLIN, DANIELE, M.		
IPC分类号	C12Q1/68 C40B30/04 G01N33/48 G01N33/50 G01N33/53 G01N33/68		
CPC分类号	C12Q1/6883 C12Q2600/118 C12Q2600/154		
代理机构(译)	MURGITROYD & COMPANY		
优先权	62/139580 2015-03-27 US		
其他公开文献	EP3274477A1		
外部链接	Espacenet		

摘要(译)

本文公开了一种用于确定后代受试者中的自闭症谱系障碍（ASD）的风险的方法。该方法包括分析含有来自预期父本亲本的精子的样品中的DNA甲基化状态，其中甲基化模式不同于与ASD无关的样品中发现的模式，表明后代患有ASD的风险。