US 20120108444A1

(19) **United States**

(12) **Patent Application Publication**

Philibert et al.

(10) **Pub. No.: US 2012/0108444 A1**

(43) **Pub. Date: May 3, 2012**

(54) **COMPOSITIONS AND METHODS FOR DETECTING PREDISPOSITION TO A SUBSTANCE USE DISORDER**

(76) Inventors: **Robert Philibert**, Iowa City, IA (US); **Anup Madan**, Bellevue, WA (US)

(21) Appl. No.: **13/284,425**

(22) Filed: **Oct. 28, 2011**

Related U.S. Application Data

(63) Continuation of application No. PCT/US2010/032815, filed on Apr. 28, 2010.

(60) Provisional application No. 61/173,274, filed on Apr. 28, 2009.

Publication Classification

(51) **Int. Cl.**

C40B 20/08	(2006.01)
C08G 83/00	(2006.01)
C12Q 1/68	(2006.01)
G01N 33/53	(2006.01)
C07H 21/04	(2006.01)
C40B 40/06	(2006.01)

(52) **U.S. Cl.** **506/6**; 536/24.31; 506/16; 435/6.11; 436/501; 525/54.2

(57) **ABSTRACT**

The present invention provides screening kits, compositions, and diagnostic methods for determining whether a subject has a predisposition to, or likelihood of having, a substance use disorder by determining a nucleic acid methylation profile from a biological sample from the subject, wherein a given profile indicates that the subject has a predisposition to a substance use disorder.

Figure 1

```
43398775            TAGTGAGGGC TGGAGGCTGC GCAGACCTCG ACGGGCCCTA CATGACGTCA
43398825            CAAAGGGGCC AGACCAAGTG GGGCAGCACC CTGCGACCCT GCGATCCTGC
43398875            CTGGCTCAGC CGCCTTCATA TATCTGCTTC CTTAAGTCCA CTCTTGCCCA
43398925            GATAGCTTTC AGTTAAAACT AAAGAATGAA AGCACTAGGT TGAGAGCCCA
43398975 CpG1-6     CGCGGCTACA CCCACGTCTA CTCCCCCACT CTCGCCAGGC AACCGCGCCC
43399025 CpG7-11    CCCGCCTGCA GTGGCATCGT CCGGCCACGC CCAGTGGCAG GGTTTCCAGC
43399075 CpG12-15   GCGAGCCTGC AGGCAGGCCG GGAAGGCGGA GCCAGGCCGG CCTAGAGTCA
43399125 CpG16-18   CTTCTCCCCG CCCCTGACTG GGCCGGGAGC CCGGGGCTGG TCTCTAAGAG
43399175            TGGGTACCGA GAACAGCCTG ACCGTGGAGA AGGGCTGCGG GAAGCAGAAC
43399225            ACCGCCCCCA GCGCCCAGCG TGCTCCAGAA ACATGAGCAC AAACGCCTCA
43399275            GCCTCCTTCC CCGGCGGCAC CGGCACCGGC ACCAGTACCC GCACCAGTAC
43399325            CGGCACCGGC ACCAGTACCC GCACCAGTAC CGGCACCGGC ACCAGTACCC
43399375            GCACCAGTAC CGGCACCGGC ACCGAGCGCA AGGCGGAGGG CCCGCCCGAA
43399425 CpG19-21   GCCGGGGGCA CAACTGCCCA GGTCCCGAAC CCGGACTCCA GCTTGGACGA
43399475 CpG22-26   CACCTCCTAC AGCCTGTCCG AATGGAGCGT CCGTTCTGAG TGGCGGTCCG
43399525 CpG27-30   TCTCGGATCC GCTAGCCAGT TCCCAGTGGA GCACGTCCTC AACTGCCGAG
43399575 CpG31-32   GCCGCCTCCT GGAGCTCCAG CATACACTCC CCAATCAGCA CTACCGGTCT
43399625 CpG33-36   TAGCGAGAGT ACTGACTCCG ACTCCAAGAG TGGCCTCCGG GGTTTCAGCG
43399675 CpG37-39   CTTACAACCC GAGCAGTCGG ATCCCCAAGT CTACCACCAG CTCGAACTCC
43399725 CpG40-42   TCCGATGGGG CCGTCACAGC CTCCAATCAG GACACCGGCA TTCCCTGGGT
43399775 CpG43-45   ATTAGTAACA GGACCTACCC CGCCCGTAAA CTCCCCCGTA GAGTCATTGC
43399825 CpG46      AAGGGTCTGC CTTCTCCTCA GGGTTCAGCA CCCCACGGGG TTTGGTAAAA
43399875 CpG47-49   GGACCGACCC TGCCCCCGGA TTCCAACCTG ACCTCAGTGT CCGACTACAC
43399925 CpG50-51   TTGGATATTT GTACGGGGAC CTCCTATACC CAATGACCTT TCGCAAGTGT
43399975 CpG52      CAATACAAGC ACCTCCTACA CCCAGTAACA CCCCCGAGTG TCAGTACAAG
43400025 CpG53      GGTCTGCCGC ATCCTCAGTG TCCAGCTTCC CCTGGGGTTT GGTACCAGGA
43400075 CpG54      CCACCTCTAC CCAATAACAT TTCCCCAGTG TCGCCACAAG CACCTCCTGC
43400125 CpG55      ACCCCATAAC ATCCCCCCAG TGTCAAGGCA GGCGTCTACC CCCACCTCAG
43400175 CpG56-57   TGCCTGACAC TCCGCGGGGT TCAATACAAG AACCTCCTGC ACCCAGTAAT
43400225 CpG58-60   CCTTTCCAGC TGCCGACACA AGGACATTCT AAACCTAATA ACTCTCGCCG
43400275 CpG61-63   AGTGTCAGTA CAAGGGTCCG CCCCGCTCTC AGTGCCCAGC TCCCCCCGGG
43400325 CpG64-66   TATCAGCTGA AACATCAGCT CCGCCCCTgg gCGctccCGg agtatcagca
43400375 CpG67-70   aaagggttCG cccCGcccac agtgccCGgc tccccCGgg tatcaaaaga
43400425 CpG71-74   aggatCGgct cCGccccCGg gctcccCGgg ggagttgata gaagggtcct
43400475 CpG75-77   tcccaccctt tgcCGtcccc actcctgtgc ctaCGaccca ggagCGtgtc
43400525 CpG78-80   agccaaagca tggagaatca agagaaggCG agtatCGCGg gccacatgtt
43400575 CpG81-86   CGaCGtagtC GtgatCGgag gtggcatttc agGTCAGTGT GGACCGTAGC
43400625            GGTGGCCTGG GGGACCCTGG CCAGTGAGGG GTAGGGGAAC CTACAGTAGC
43400675 CpG87      TCTTGTGGTG TTTGGGGGTC TCTCATGCAT GCGAGAGTGT AGTGTAGCCA
43400725 CpG88      TGGCTTGGCC CCATATCCTG CGAGGTAGGA GTGGGGGTTG TGCCAGTTTT
43400775            GCTGGTGGTG TGACTGGGGG AGGCAGACAC AATAATTTTA CTACTACTAC
                    (SEQ ID NO: 9)
```

Figure 2

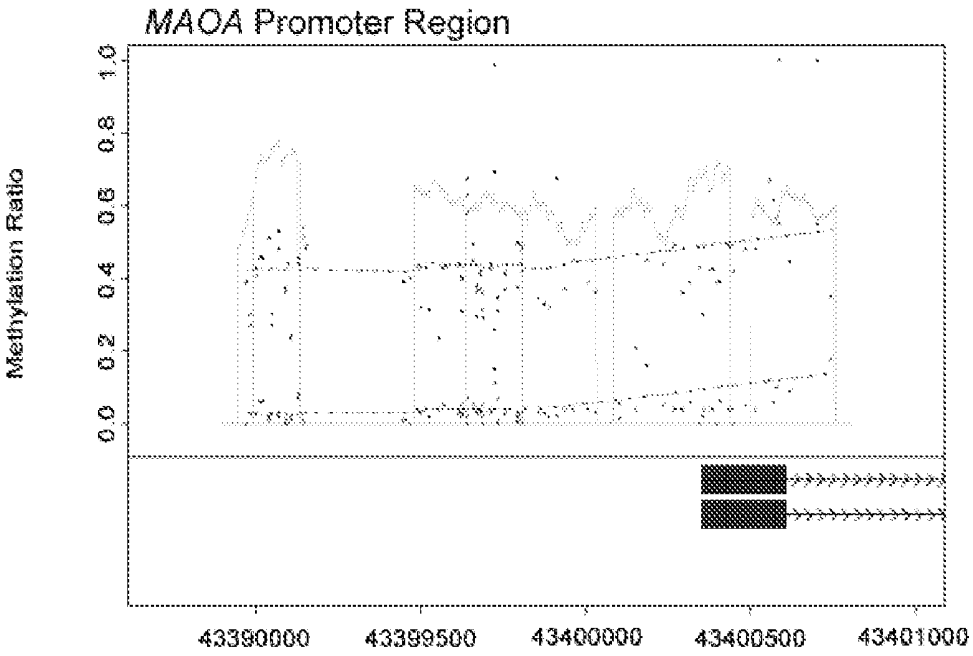
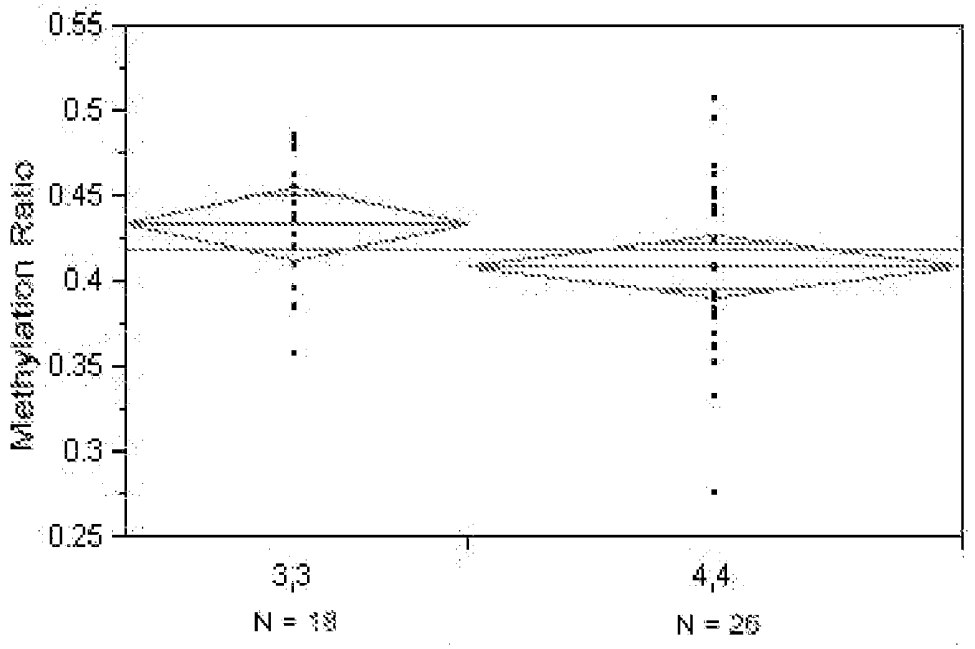
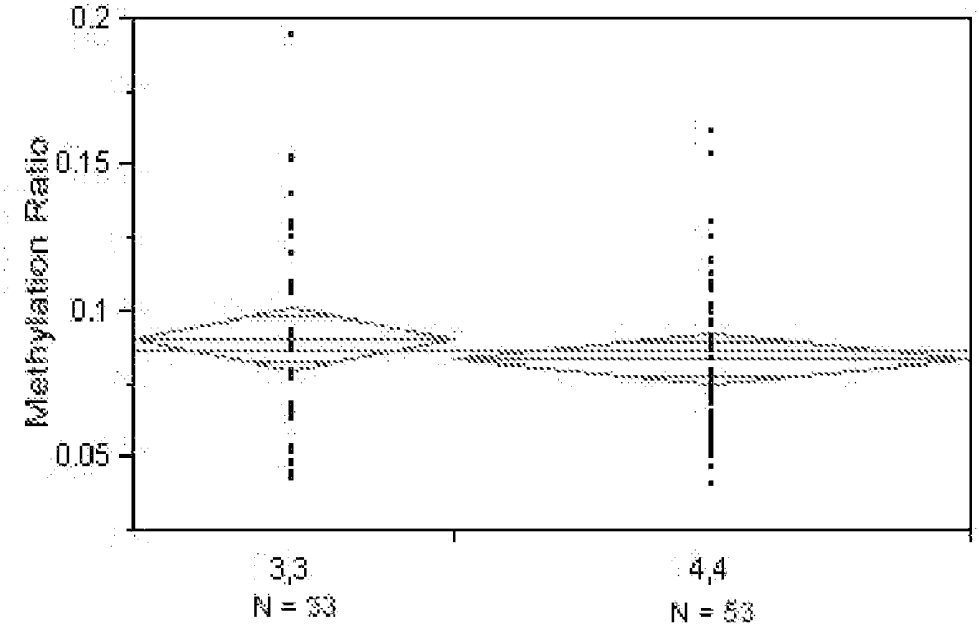

MAOA Promoter Region
Methylation Ratio
0.0
0.2
0.4
0.6
0.8
1.0
43399000
43399500
43400000
43400500
43401000

Figure 3


0.55
0.5
0.45
0.4
0.35
0.3
0.25
Methylation Ratio
3,3
N = 19
4,4
N = 26


0.2
0.15
0.1
0.05
Methylation Ratio
3,3
N = 33
4,4
N = 53

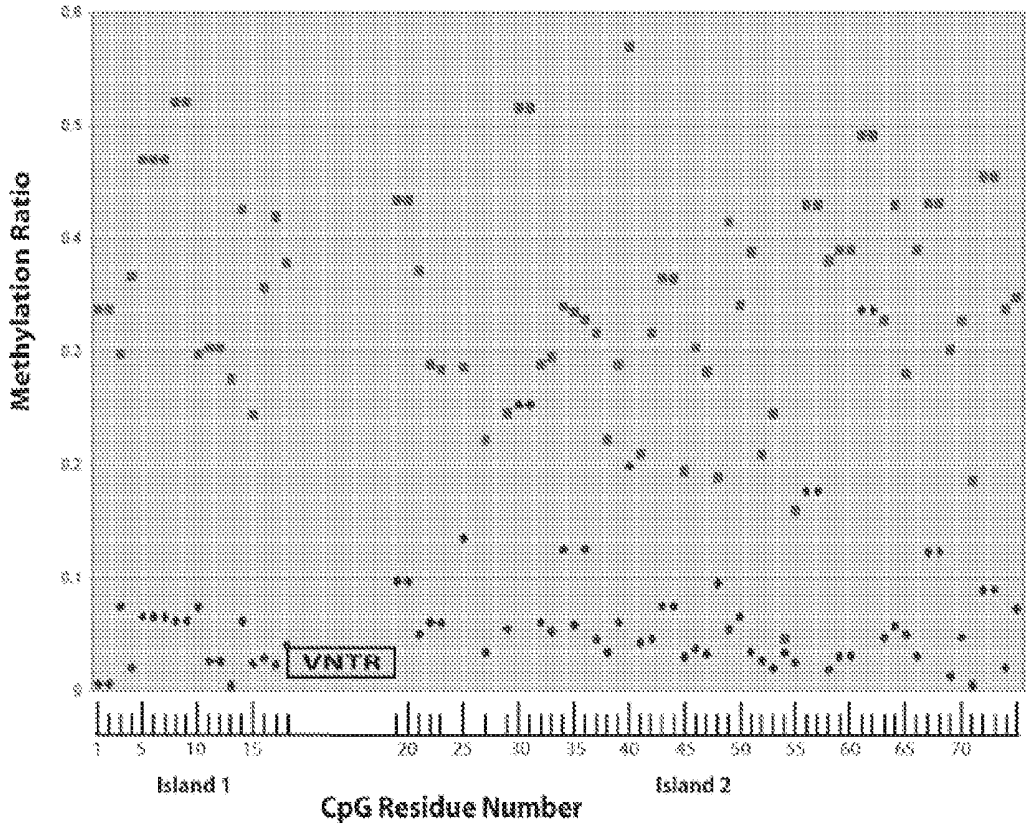

Methylation Ratio
VNTR
1
5
10
15
20
25
30
35
40
45
50
55
60
65
70
Island 1
Island 2
CpG Residue Number

Figure 4

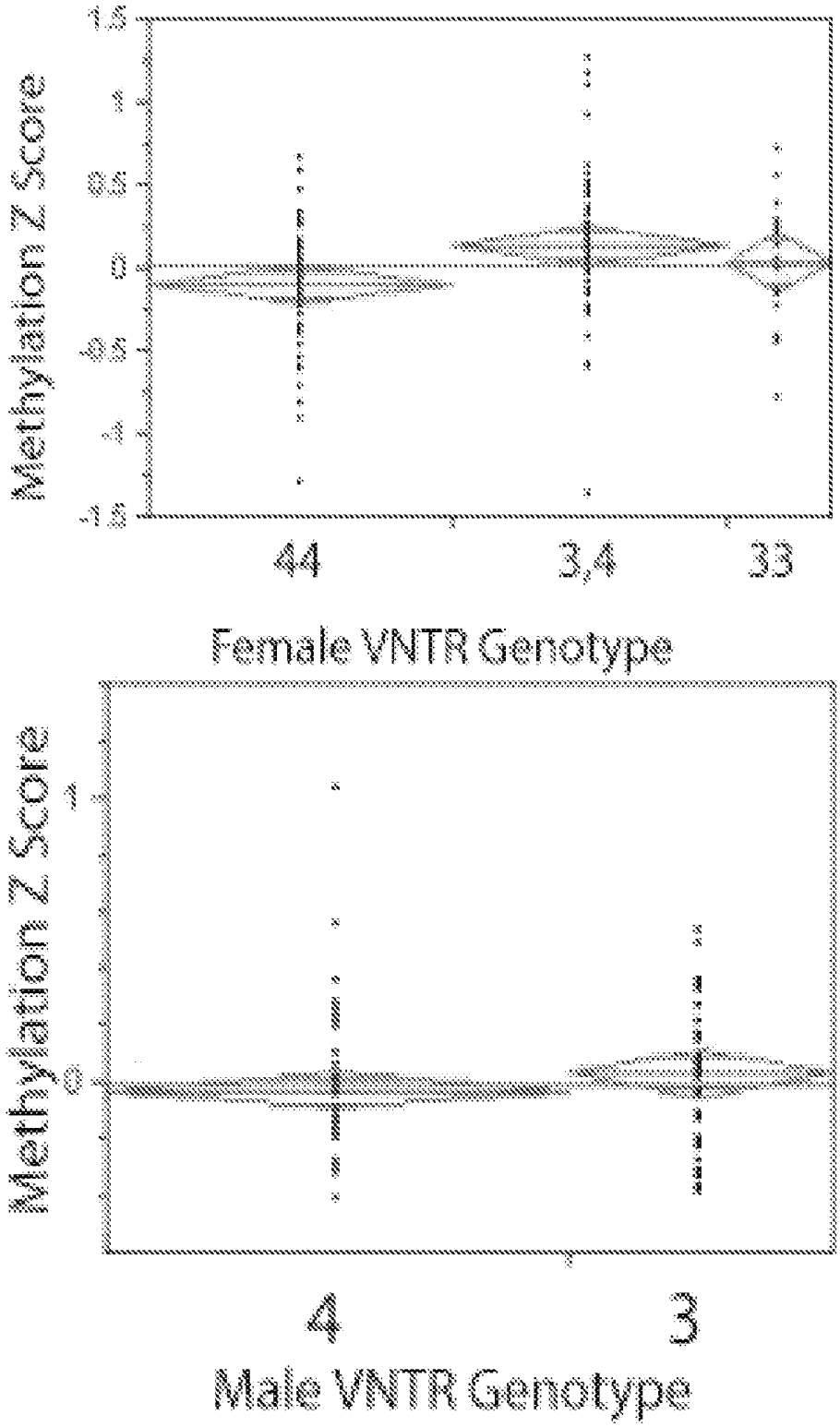

Methylation Z Score
1.5
1
0.5
0
-0.5
-1
-1.5
44
3,4
33
Female VNTR Genotype
Methylation Z Score
1
0
4
3
Male VNTR Genotype

Figure 5

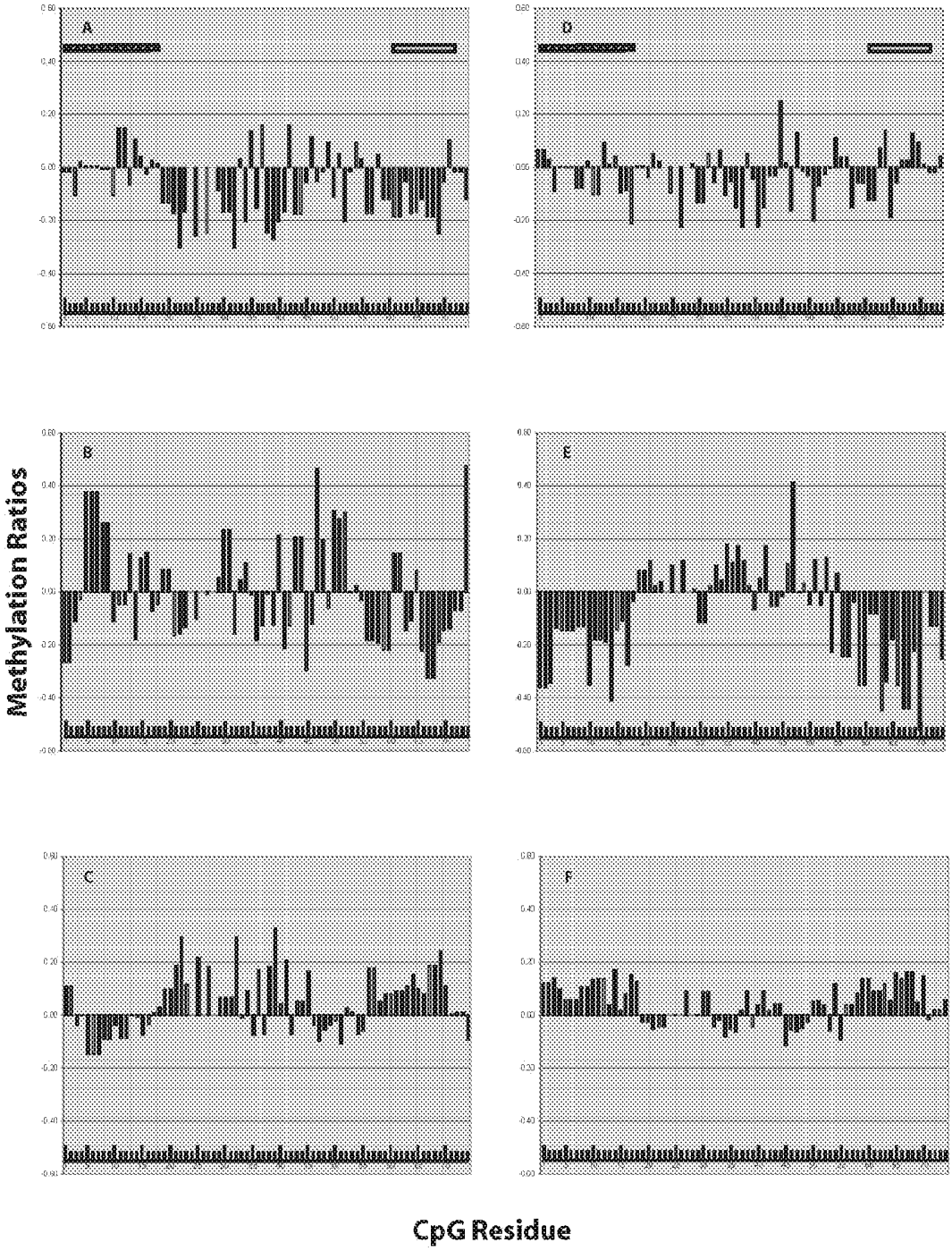


Figure 6

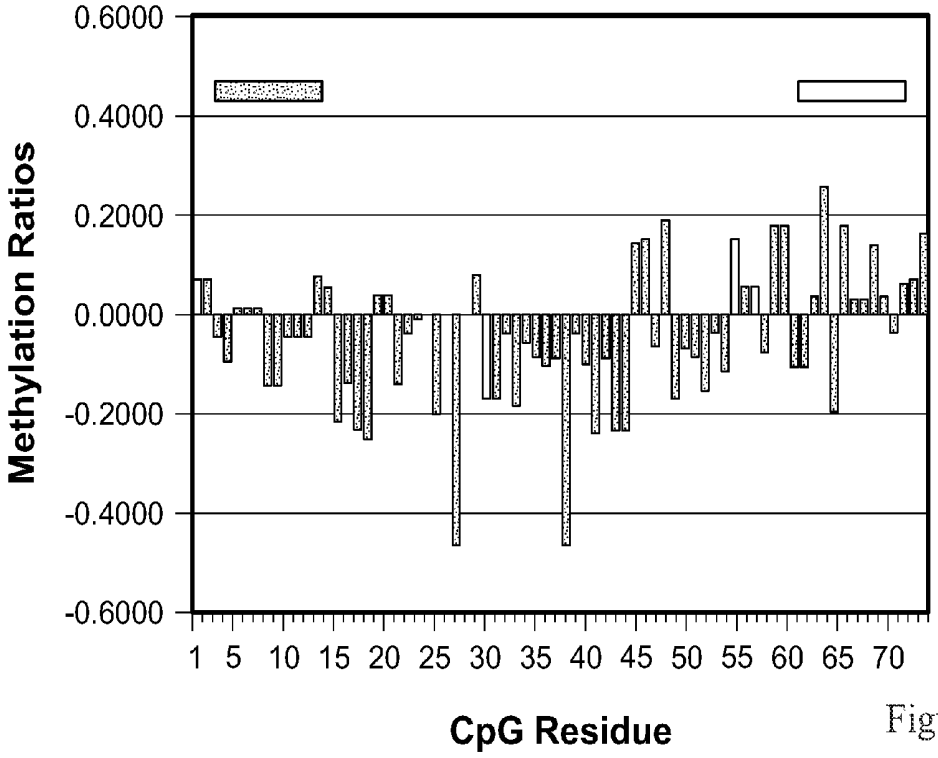


Figure 7A

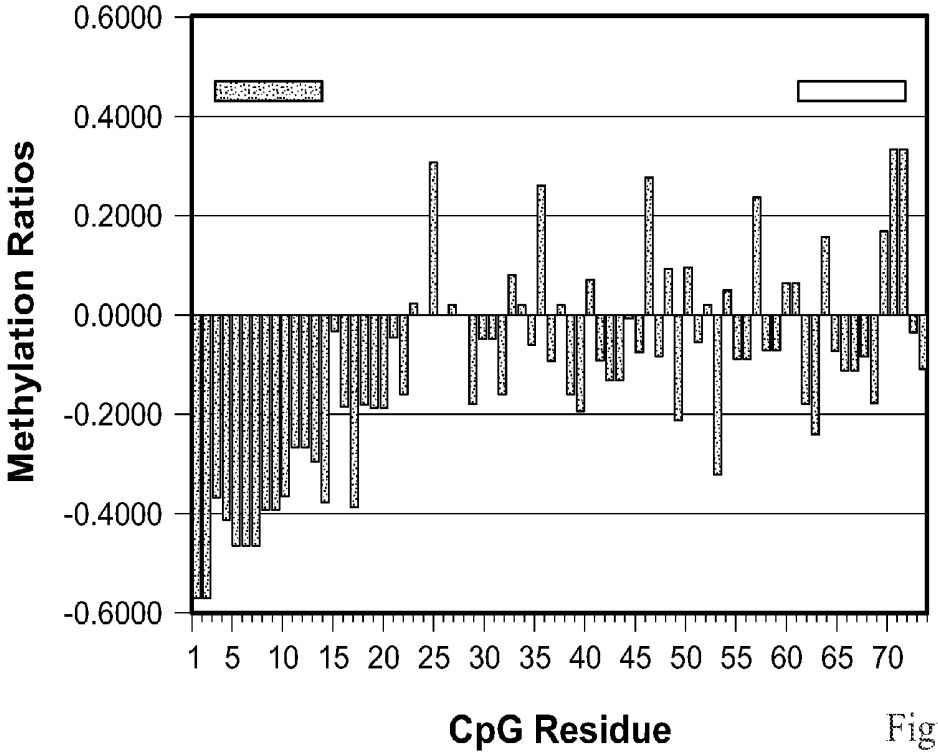


Figure 7B

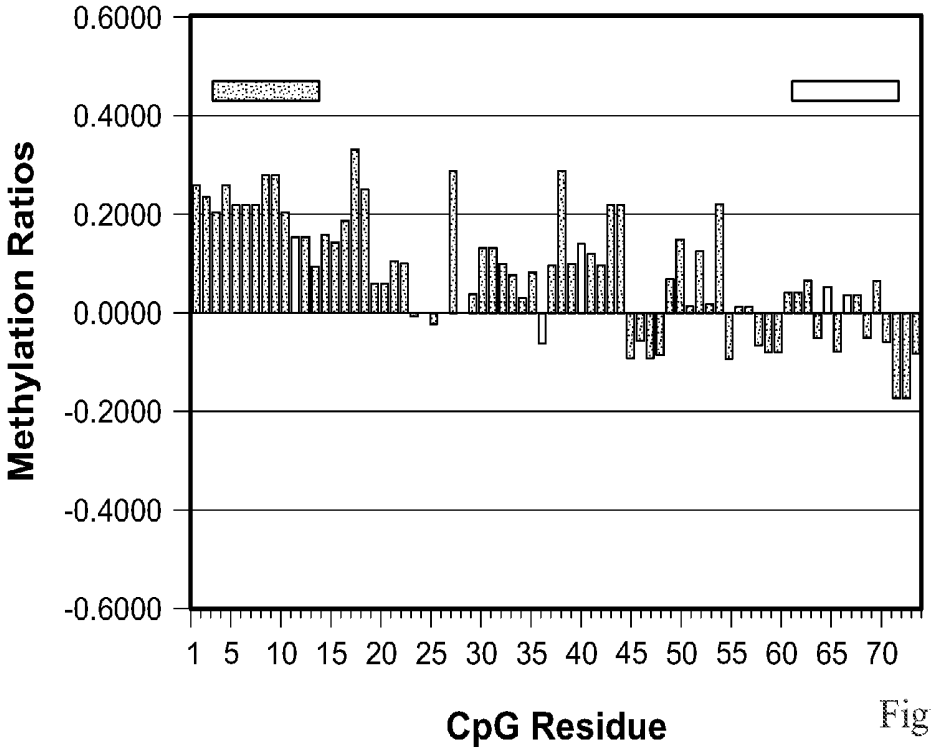


Figure 7C

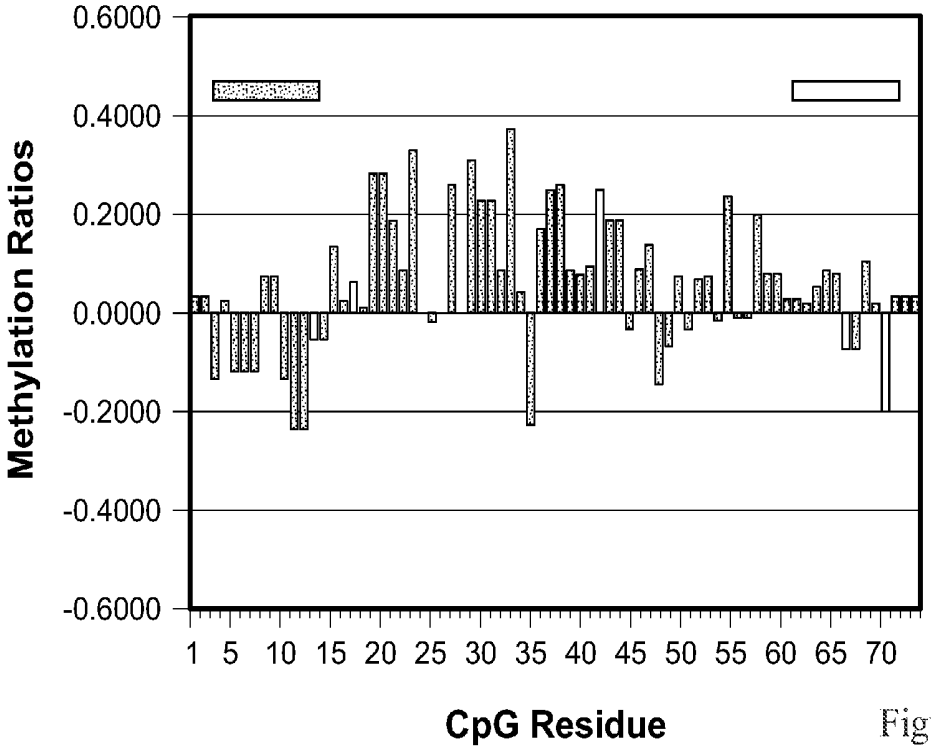


Figure 7D

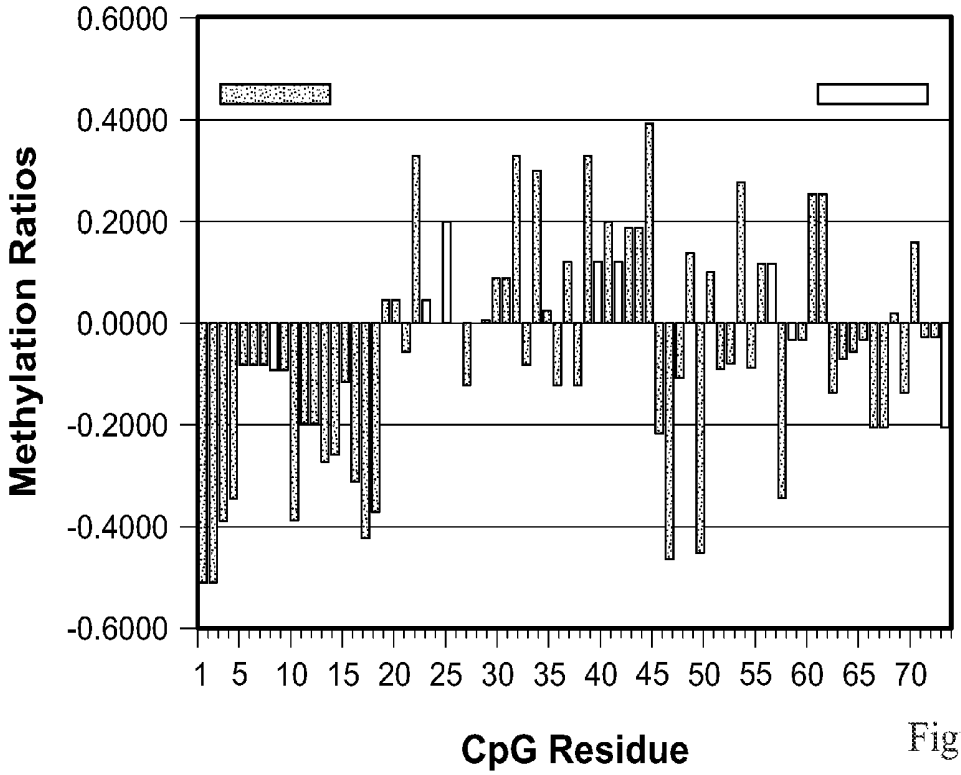


Figure 7E

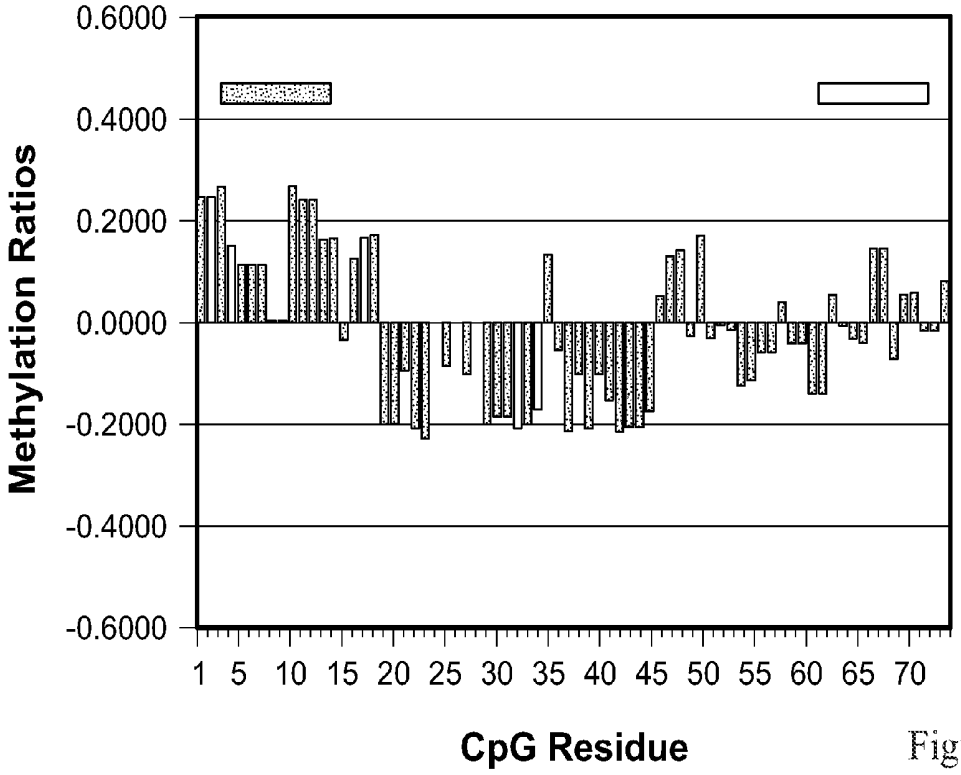


Figure 7F

Figure 8. Sequence and Probe Placement for the *AX2R* Promoter

ggttgaggctgcagtgagctatgcttgtgccactgcactccagcctgggggacacagccagaccctgtctcaaaaaaagtgaaaaaaaaaaaaaag
ccaaaaaatactgtggagtggccctttcttcaagcatctgcttgctttctctaacaccactcattctattgcccctactgagcttgaaatgataatgctttcttc
aggtgccaggtctggagtgctggtgcacctatctcaaaaCGctgtctcaaaactccaacagggagcacctaaCGgtactgggtgcaacattgcta CpG 1-2
gcaCGgagcaaacagcagtccagtaacctggaacaacaggctctgCGaaaccaaggactctgacaaagaaaaattgccaattccaaacatagc CpG 3-4
ctgttttagagaaatgaatggCGttgtcatCGaaaaaacacagactCGattgtgacagaaatacCGccacaaaCGcaggtacagggacagc CpG 5-9
CGacacCGagaaccaagggaagCGgctgagagctgCGcctccaCGgaataactgccagcCGgcacagtgCGagtgagaaacCG CpG 10-17
gccactccatgaaaCGaccagtactgccaCGgaaaagaatcCGaCGtCGccaacaagCGgtgctaccaggagaaaCGcctgcttttg CpG 18-24
aagaaaacagccaggaaCGCGactgaaagacacttgctcccaggaagaattggcatttgttccaaaacacagctggataaacCGagaaccttC CpG 25-28
Ggagtggttgcac CGaaaCGgggtcacccagcacctcagCGtcctgggctctagcaagcctcacagaagcaacCGgcagtgctaacacC CpG 29-33
Gaggagcacctagag CGgcaaaactagcagtaatgccatCGaCGaaaggccagttaggccaaaagaatagaatatttagttcCGggaattac CpG 33-35
aggccagCGcaaaccagacagcataaagctgagggtcagcaaaacaaaaattaggaacaatttttttttaaagggcaagttagctgaaaaacacaC CpG 36-37
Gcacacacaataaaaacaatacatttgggaagattcatcaaatgaaaattcaaaactaagcaaaacatggaaaaatggactctacaaagaagaaaatgc
(SEQ ID NO:21)

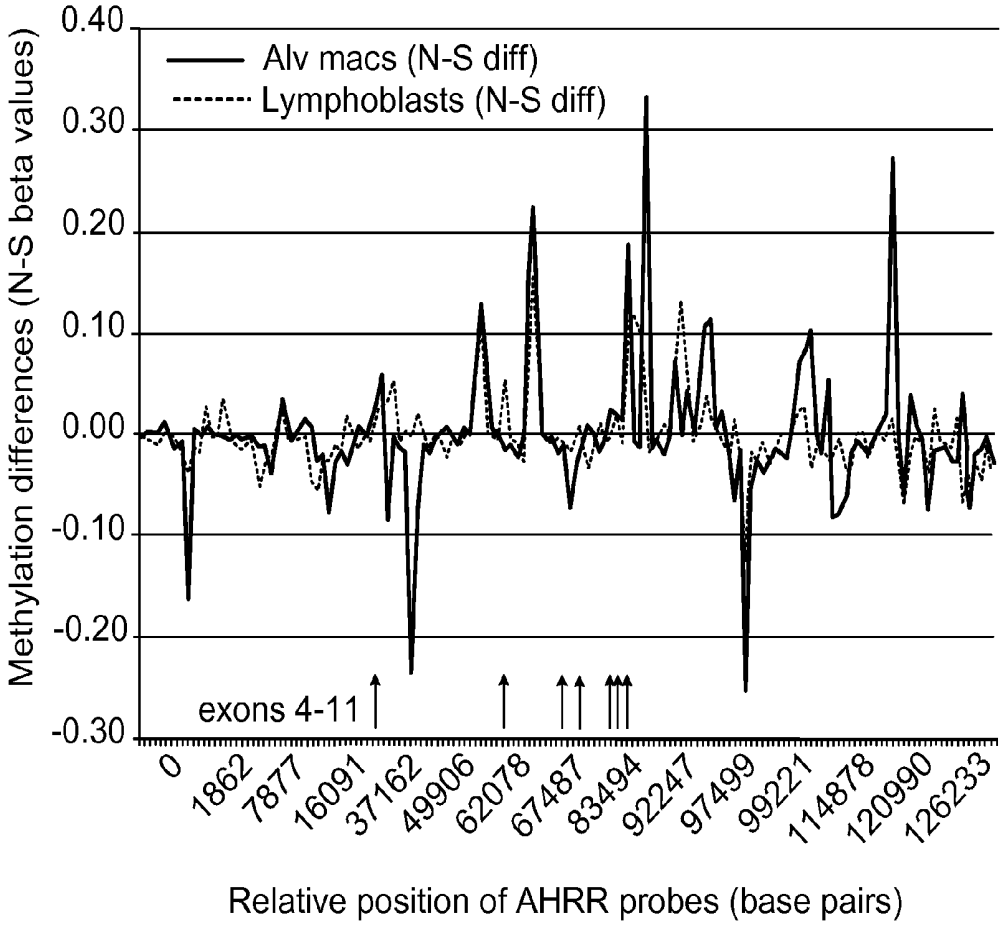

0.40
0.30
0.20
0.10
0.00
-0.10
-0.20
-0.30
Methylation differences (N-S beta values)
Alv macs (N-S diff)
Lymphoblasts (N-S diff)
exons 4-11
0
1862
7877
16091
37162
49906
62078
67487
83494
92247
97499
99221
114878
120990
126233
Relative position of AHRR probes (base pairs)

Figure 9

COMPOSITIONS AND METHODS FOR DETECTING PREDISPOSITION TO A SUBSTANCE USE DISORDER

RELATED APPLICATION

[0001] This application claims the benefit of priority under 35 U.S.C. §119 to PCT/US2010/032815, filed Apr. 28, 2010, which claims the benefit of priority under 35 U.S.C. §119(e) to U.S. Application No. 61/173,274, filed Apr. 28, 2009. The entirety of these disclosures is incorporated herein by reference.

STATEMENT OF GOVERNMENT SUPPORT

[0002] Work related to this invention was funded by the U.S. government (NIH Grants DA015789, DA010923, DA02173603, MH080898, and P30DA027827). The government has certain rights in this patent.

BACKGROUND

[0003] Substance use disorders cause serious problems, both for the affected individuals and for society in general. Despite intensive research, however, a reliable laboratory test for diagnosing a patient as having, or for being at risk for developing, such conditions has not been developed. Such diagnoses are still generally made clinically, on the basis of observed behavior. Given the difficulties of defining normal experience and behavior and the lack of reliable objective indicators, it is not surprising that to date systems of diagnosis in psychiatry have been less than satisfactory. A reliable laboratory test would be of practical value in everyday clinical practice, for example, in assisting doctors in prescribing the appropriate treatment for their patients. Thus, methods of identifying subjects that have, or are at risk for developing, substance use disorders are needed.

SUMMARY OF CERTAIN EMBODIMENTS OF THE INVENTION

[0004] The present invention provides a screening kit for determining whether a human subject has the likelihood of using, abusing or being dependent upon a substance comprising: (a) a solid substrate, at least one probe specific for methylation status of a CpG dinucleotide repeat motif expressed by a peripheral blood cell or its derivative, wherein the methylation status of the CpG dinucleotide is associated with nicotine use, abuse or dependence; and/or (b) a solid substrate, at least one probe specific for methylation status a CpG dinucleotide repeat motif expressed by a peripheral blood cell or its derivative, wherein the methylation status of the CpG dinucleotide is associated with alcohol use, abuse or dependence; and/or (c) a solid substrate, at least one probe specific for methylation status a CpG dinucleotide repeat motif expressed by a peripheral blood cell or its derivative, wherein the methylation status of the CpG dinucleotide is associated with cannabis use, abuse or dependence. As used herein, the term "methylation status" means the determination whether a certain target DNA, such as a CpG dinucleotide, is methylated. As used herein the term "CpG dinucleotide repeat motif" means a series of two or more CpG dinucleotides positioned in a DNA sequence.

[0005] In certain embodiments of the present invention, the substance is nicotine and the CpG dinucleotide repeat motif is located in a gene from Table 5 or Table 9 or Appendix C. In certain embodiments of the present invention, the substance is alcohol and the CpG dinucleotide repeat motif is located in a gene from Table 6 or Table 13 or Appendix E. In certain embodiments of the present invention, the substance is cannabis and the CpG dinucleotide repeat motif is located in a gene from Table 7 or Table 14.

[0006] The present invention also provides a screening kit for determining whether a subject has a predisposition to, or likelihood of having, a substance use disorder including at least one probe specific for a methylated monoamine oxidase A (MAOA) or monoamine oxidase B (MAOB) locus or a aryl hydrocarbon receptor repressor (AHRR) locus in a peripheral blood cell, wherein the methylation of MAOA or AHRR is associated with a substance use disorder. In certain embodiments, the kit further includes a solid substrate, wherein each probe is bound onto the substrate in a distinct spot. In certain embodiments, the substance use disorder is nicotine dependence. In certain embodiments, the probe detects methylation at CpG residue 18, 42, 48, 52, 64, 65, 66, 67, 68, 69, and/or 77 (for MAOA) and at CpG residues in the AHRR gene. In certain embodiments, the substance use disorder is alcohol dependence. In certain embodiments, the probe detects methylation at CpG residue 27, 38, 41 and/or 48 (for MAOA). In certain embodiments, the substance use disorder is cannabis dependence. In certain embodiments, the subject is female and the probe detects methylation at CpG residue 69 and/or 88 (for MAOA). In certain embodiments, the subject is male and the probe detects methylation at CpG residue 11-12, 13, 64, 69, 72 and/or 73 of MAOA. In certain embodiments, the substrate is a polymer, glass, semiconductor, paper, metal, gel or hydrogel. In certain embodiments, the kit further includes at least one control probe, wherein the at least one control probe is bound onto the substrate in a distinct spot. In certain embodiments, the solid substrate is a microarray or microfluidics card. In certain embodiments, the probe is an oligonucleotide probe or a nucleic acid derivative probe.

[0007] The present invention provides a screening kit that uses bisulfite treated DNA for determining whether a subject has the likelihood of using, abusing or being dependent upon a substance comprising: (a) a single base pair extension probe, with at least one probe specific for methylation status of a CpG dinucleotide repeat motif expressed by a peripheral blood cell or its derivative, wherein the methylation status of the CpG dinucleotide is associated with nicotine use, abuse or dependence; and/or (b) a single base pair extension probe, at least one probe specific for methylation status of a CpG dinucleotide repeat motif expressed by a peripheral blood cell or its derivative, wherein the methylation status of the CpG dinucleotide is associated with alcohol use, abuse or dependence; and/or (c) a single base pair extension probe, at least one probe specific for methylation status of a CpG dinucleotide repeat motif expressed by a peripheral blood cell or its derivative, wherein the methylation status of the CpG dinucleotide is associated with cannabis use, abuse or dependence. As used herein, a "single base pair extension probe" is a nucleic acid that selectively recognizes a single nucleotide polymorphism (i.e., either the A or the G of an A/G polymorphism). Generally, these probes take the form of a DNA primer (e.g., as in PCR primers) that are modified so that incorporation of the primer releases a fluorophore. One example of this is a Taqman® probe that uses the 5' exonuclease activity of the enzyme Taq Polymerase for measuring the amount of target sequences in the samples. TaqMan® probes consist of a 18-22 bp oligonucleotide probe, which is labeled with a reporter fluorophore at the 5' end, and a

quencher fluorophore at the 3' end. Incorporation of the probe molecule into a PCR chain (which occurs because the probe set is contained in a mixture of PCR primers) liberates the reporter fluorophore from the effects of the quencher. The primer must be able to recognize the target binding site. Some primer extension probes can be "activated" directly by DNA polymerase without a full PCR extension cycle.

[0008] The present invention provides a screening kit that uses bisulfite treated DNA for determining whether a subject has the likelihood of having a substance use disorder or substance use syndrome comprising: (a) a nucleic acid primer, with at least one primer specific for methylation status of a CpG dinucleotide repeat motif region contained by a peripheral blood cell or its derivative, wherein the methylation status of the CpG dinucleotide is associated with nicotine use, abuse or dependence; and/or (b) a nucleic acid primer, at least one primer specific for methylation status of a CpG dinucleotide repeat motif region contained by a peripheral blood cell or its derivative, wherein the methylation status of the CpG dinucleotide is associated with alcohol use, abuse or dependence; and/or (c) a nucleic acid primer, at least one primer specific for methylation status of a CpG dinucleotide repeat motif region contained by a peripheral blood cell or its derivative, wherein the methylation status of the CpG dinucleotide is associated with cannabis use, abuse or dependence. In certain embodiments, the kit may contain a number of primers that is any integer between 1 and 10,000, such as 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, . . . 9997, 9998, 9999, 10,000. As used herein, the term "nucleic acid primer" encompasses both DNA and RNA primers.

[0009] The present invention provides a diagnostic method using bisulfite treated DNA for determining whether a subject has the likelihood of having a substance use disorder or substance use syndrome comprising: (a) determining methylation status of a CpG dinucleotide repeat motif region in a peripheral blood cell or its derivative, wherein the methylation status of the CpG dinucleotide is associated with nicotine use, abuse or dependence; and/or (b) determining methylation status of a CpG dinucleotide repeat motif region in a peripheral blood cell or its derivative, wherein the methylation status of the CpG dinucleotide is associated with alcohol use, abuse or dependence; and/or (c) determining methylation status of a CpG dinucleotide repeat motif region in a peripheral blood cell or its derivative, wherein the methylation status of the CpG dinucleotide is associated with cannabis use, abuse or dependence. In certain embodiments, the method determines the methylation status of a plurality of CpG dinucleotide repeat motif regions. Such a plurality may be any integer between 1 and 10,000, such as at least 100.

[0010] The present invention provides a diagnostic method for determining whether a subject has a predisposition to, or likelihood of having, a substance use disorder, by determining a nucleic acid methylation profile from a single type of peripheral blood cell or blood cell derivative from the subject, the method comprising: (a) obtaining a profile associated with the sample, wherein the profile comprises quantitative data for methylation of a monoamine oxidase A (MAOA) locus or for methylation of a AHRR locus in the blood cell; (b) inputting the data into an analytical process that uses the data to classify the sample, wherein the classification is a "substance use disorder" classification or a "healthy" classification; and (c) classifying the sample according to the output of the process.

[0011] In certain embodiments of the present invention, the blood cell is a lymphocyte, such as a monocyte, a basophil, an eosinophil, and/or a neutrophil. In certain embodiments, the blood cell type is a mixture of peripheral white blood cells. In certain embodiments, the peripheral blood cell has been transformed into a cell line.

[0012] In certain embodiments, the analytical process comprises comparing the obtained profile with a reference profile. In certain embodiments, the reference profile comprises data obtained from one or more healthy control subjects, or comprises data obtained from one or more subjects diagnosed with a substance use disorder. In certain embodiments, the method further comprises obtaining a statistical measure of a similarity of the obtained profile to the reference profile. In certain embodiments, the blood cell or blood cell derivative is a peripheral blood cell. In certain embodiments, the profile is obtained by sequencing of methylated DNA, such as by digital sequencing.

[0013] The present invention provides a diagnostic method for determining whether a subject has a predisposition to, or likelihood of having, a substance use disorder, by determining a nucleic acid methylation profile from a single type of blood cell or blood cell derivative from the subject, the method involves: (a) obtaining a profile associated with the sample, wherein the profile determines quantitative data for methylation of a monoamine oxidase A (MAOA) locus or for methylation of a AHRR locus in the blood cell; (b) inputting the data into an analytical process that uses the data to classify the sample, wherein the classification is a "substance use disorder" classification or a "healthy" classification; and (c) classifying the sample according to the output of the process. In certain embodiments, the analytical process involves comparing the obtained profile with a reference profile. In certain embodiments, the reference profile provides data obtained from one or more healthy control subjects, or provides data obtained from one or more subjects diagnosed with a substance use disorder. In certain embodiments, the method further involves obtaining a statistical measure of a similarity of the obtained profile to the reference profile. In certain embodiments, the blood cell or blood cell derivative is a peripheral blood cell. In certain embodiments, the blood cell is a lymphocyte. In certain embodiments, the lymphocyte type is a B-lymphocyte. In certain embodiments, the B-lymphocytes have been immortalized. In certain embodiments, the blood cell type is a monocyte. In certain embodiments, the blood cells type is a basophil. In certain embodiment, the substance use disorder is nicotine dependence, alcohol dependence, or cannabis dependence.

[0014] In certain embodiments, a solid substrate may contain a number of probes that is any integer between 1 and 10,000 probes, such as 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, . . . , 9997, 9998, 9999, 10,000. In one kit, all of the probes may be physically located on a single solid substrate or on multiple substrates.

[0015] In certain embodiments, the current invention can also take the form of a PCR (polymerize chain reaction) assay. In some cases, this will take the form of real time PCR assays (RT-PCR) assays. In certain embodiments of these PCR assays, a kit may contain two primers that specifically amplify a region of a MAOA locus or a AHRR locus and gene specific probe that selectively recognizes the amplified region. Together, the primers and the gene specific probes are referred to as a primer-probe set. By measuring the amount of gene specific probe that has hybridized to an amplified seg-

ment at a given point of the PCR reaction or throughout the PCR reaction, one who is skilled in the art can infer the amount of nucleic acid originally present at the start of the reaction. In some cases, the amount of probe hybridized is measured through fluorescence spectrophotometry. The number of primer-probe sets can be any integer between 1 and 10,000 probes, such as 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, . . . 9997, 9998, 9999, 10,000. In one kit, all of the probes may be physically located in a single reaction well or in multiple reaction wells. The probes may be in dry or in liquid form. They may be used in a single reaction or in a series of reactions. In certain embodiments, the probe is an oligonucleotide probe. In certain embodiments, the probe is a nucleic acid derivative probe.

[0016] The term "substrate" refers to any solid support to which the probes may be attached. The substrate material may be modified, covalently or otherwise, with coatings or functional groups to facilitate binding of probes. Suitable substrate materials include polymers, glasses, semiconductors, papers, metals, gels and hydrogels among others. Substrates may have any physical shape or size, e.g., plates, strips, or microparticles.

[0017] The term "spot" refers to a distinct location on a substrate to which probes of known sequence or sequences are attached. A spot may be an area on a planar substrate, or it may be, for example, a microparticle distinguishable from other microparticles.

[0018] The term "bound" means affixed to the solid substrate. A spot is "bound" to the solid substrate when it is affixed in a particular location on the substrate for purposes of the screening assay.

[0019] In certain embodiments of the kit of the present invention, the substrate is a polymer, glass, semiconductor, paper, metal, gel or hydrogel. In certain embodiments of the present invention, the kit further includes a solid substrate and at least one control probe, wherein the at least one control probe is bound onto the substrate in a distinct spot.

[0020] In certain embodiments of the present invention, the solid substrate is a microarray. An "array" or "microarray" is used synonymously herein to refer to a plurality of probes attached to one or more distinguishable spots on a substrate. A microarray may include a single substrate or a plurality of substrates, for example a plurality of beads or microspheres. A "copy" of a microarray contains the same types and arrangements of probes.

[0021] The present invention also provides a composition for determining whether a subject has a predisposition to, or likelihood of having, a substance use disorder by determining a nucleic acid methylation profile from a single type of blood cell or blood cell derivative from the subject, the method including obtaining a profile associated with the sample, wherein the profile includes quantitative data for MAOA; (b) inputting the data into an analytical process that uses the data to classify the sample, wherein the classification is a "substance use disorder" classification or a "healthy" classification; and (c) classifying the sample according to the output of the process. In certain embodiments, a solid substrate may contain a number of probes that is any integer between 1 and 10,000 probes, such as 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, . . . 9997, 9998, 9999, 10,000.

[0022] As used herein, the term "healthy" means that a subject does not manifest a particular condition, and is no more likely that at random to be susceptible to a particular condition.

[0023] The present invention also provides a composition for determining whether a subject has a predisposition to, or likelihood of having nicotine dependence, alcohol dependence or cannabis dependence including (a) a solid substrate; (b) at least one probe specific for a methylated MAOA gene or AHRR gene associated with nicotine dependence, alcohol dependence or cannabis dependence wherein each probe is bound onto the substrate in a distinct spot. In certain embodiments, a solid substrate may contain a number of probes that is any integer between 1 and 10,000 probes, such as 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, . . . 9997, 9998, 9999, 10,000.

[0024] In addition to the specific biomarker sequences identified in this application by name, accession number, or sequence, the invention also contemplates use of biomarker variants that are at least 90% or at least 95% or at least 97% identical to the exemplified sequences and that are now known or later discover and that have utility for the methods of the invention. These variants may represent polymorphisms, splice variants, mutations, and the like. Various techniques and reagents find use in the diagnostic methods of the present invention. In one embodiment of the invention, blood samples, or samples derived from blood, e.g. plasma, circulating, etc. are assayed for the presence of polypeptides. Typically a blood sample is drawn, and a derivative product, such as plasma or serum, is tested. Such polypeptides may be detected through specific binding members. The use of antibodies for this purpose is of particular interest. Various formats find use for such assays, including antibody arrays; ELISA and RIA formats; binding of labeled antibodies in suspension/solution and detection by flow cytometry, mass spectroscopy, and the like. Detection may utilize one or a panel of antibodies, preferably a panel of antibodies in an array format. Expression signatures typically utilize a detection method coupled with analysis of the results to determine if there is a statistically significant match with a disease signature.

[0025] The present invention also provides a composition for determining whether a subject has a predisposition to, or likelihood of having nicotine dependence, alcohol dependence or cannabis dependence including a PCR or RT-PCR assay kit containing at least one primer-probe set specific for a methylated MAOA nucleic acid or a methylated AHRR nucleic acid.

[0026] The present invention also provides a diagnostic method for determining whether a subject has a predisposition to, or likelihood of having, a substance use disorder by determining a nucleic acid methylation profile from a single type of blood cell or a blood cell derivative from the subject, the method involves (a) obtaining a profile associated with the sample, wherein the profile comprises quantitative data for at least one methylated MAOA nucleic acid or for at least one methylated AHRR nucleic acid; (b) inputting the data into an analytical process that uses the data to classify the sample, wherein the classification is a "substance use disorder" classification or a "healthy" classification; and (c) classifying the sample according to the output of the process. In certain embodiments, the analytical process comprises comparing the obtained profile with a pre-determined reference profile. In certain embodiments the reference profile comprises data obtained from one or more healthy control subjects, or comprises data obtained from one or more subjects diagnosed with a substance use disorder. In certain embodiments, the method further involves obtaining a statistical measure of a similarity of the obtained profile to the reference profile.

[0027] In certain embodiments the blood cell is a lymphocyte. In certain embodiments the lymphocyte type is a B-lymphocyte. In certain embodiments, the B-lymphocytes have been immortalized. In certain embodiments, the blood cell type is a monocyte. In certain embodiments, the blood cell type is a basophil.

[0028] The present invention provides a diagnostic method for determining whether a subject has a predisposition to, or likelihood of having, a substance use disorder. As used herein the term "predisposition" is defined as a tendency or susceptibility to manifest a condition. A subject is more likely than a control subject to manifest the condition. The term "substance use disorder" includes both abuse and dependence on a substance. The method involves determining a nucleic acid methylation profile from cells in a biological sample from the subject, wherein a given profile indicates that the subject has a predisposition to, or likelihood of having, a substance use disorder. The substance use disorder to be diagnosed may include nicotine dependence and/or alcohol dependence.

[0029] The present invention also provides a method for diagnosing a predisposition to, or likelihood of having, a substance use disorder, where the method involves (a) determining a nucleic acid methylation profile of MAOA or AHRR from a single type of cell from a biological sample from the subject; and (b) comparing the nucleic acid methylation profile with a nucleic acid methylation profile characteristic of the condition to determine if the patient has the a predisposition to, or likelihood of having, a substance use disorder.

[0030] The present invention further provides a method for evaluating and treating a patient experiencing a substance use disorder, where the method involves (a) obtaining a baseline laboratory profile comprising collecting blood from the patient to determine the patient's baseline nucleic acid methylation of MAOA or AHRR profile level from a single type of cell; (b) treating the patient for the substance use disorder; (c) obtaining a post-treatment laboratory profile comprising collecting blood from the patient to determine the patient's post-treatment nucleic acid methylation profile level from the same type of cell tested previously; and (d) comparing the baseline and post-treatment laboratory profile to evaluate the effectiveness of the treatment.

BRIEF DESCRIPTION OF THE FIGURES

[0031] FIG. **1**. The sequence and structure of the MAOA promoter region (SEQ ID NO:9). The first CpG island begins at by 43398975 and contains 18 CpG residues. A second CpG island begins at by 43399493 and contains 70 CpG residues. The position of each of the CpG residues is noted in the figure. The first exon of MAOA is denoted by small letters and is wholly contained within the second island. The positions of the primers used to amplify the MAOA VNTR are denoted by boxed letters. The transcription start site (TSS) is at bp43400353 between CpG residues 64 and 65.

[0032] FIG. **2**. The average methylation ratios (methyl CpG/total CpG) at each CpG residue for each sex. The by position on the X chromosome is given on the X axis and corresponds to the position of each of the residues in FIG. **1**. The average values for female subjects are depicted by blue squares, while the average values for males are depicted by red circles. The position of MAOA exon 1 is denoted by the box with the direction of transcription being indicated by the line with arrows.

[0033] FIG. **3**. The relationship of MAOA VNTR genotype to methylation in females (above) and males (below). There was a trend for association for female 3,3 homozygotes to have higher average methylation (methyl CpG/total CpG) than female 4,4 homozygotes (43.3%±3.8 vs 40.9%±5.2; $p<0.10$). There was no significant difference between males hemizygous for the 3 repeat allele as compared to those with the 4 allele although the arithmetic difference was in the same direction (9.0±3.7 vs 8.3±2.6; $p<0.32$).

[0034] FIG. **4**. The average methylation (methyl CpG/total CpG) at each CpG residue for each sex. The first island consists of 18 CpG residues while the second larger island consists of 70 residues, of which only the first 56 were analyzed in this study. Tic marks at the positions corresponding to CpG 24, 26 and 28 are missing because average methylation could not be reliably determined at those residues. The average methylation value for females at each residue is depicted by a pink square while the corresponding value for males is depicted by a blue diamond. The overall average methylation value is depicted by the value corresponding at position 75 (34.8% and 7.2% for females and males, respectively). The exact position of the transcription start site is between CpG 65 and CpG 66.

[0035] FIG. **5**. The relationship of MAOA VNTR genotype to methylation in females (above) and males (below). There was no significant difference between males with the 3R (n=35; mean Z score −0.03, non-transformed average methylation (NTWAM) is 7.1%), and those with a 4R allele (n=61; mean Z score −0.03, NTWAM is 7.2%). Female 4R homozygotes (Z=−−0.101, NTWAM 33.6%) had significantly lower methylation than 3,4 heterozygotes (Z=0.137, NTWA 36.2%; $p<0.01$). The difference between 4R and 3R homozygotes (Z=0.007, NTWAM, 34.7%) was not statistically different ($p<0.39$).

[0036] FIGS. **6**A-**6**F. Plot of average methylation Z score at each residue in LB DNA for each grouping of smoking status. CpG residues are in order from left to right. The hatched bar indicates the residues in the first promoter island. The open bar indicates the TSS region. Group A. Current daily male smokers (n=42). Group B. Males who have quit smoking (n=20). Group C. Males who have never smoked daily (n=59). Group D. Current daily female smokers (n=45). Group E. Females who have quit (n=27). Group F. Females who have never smoked daily (n=83).

[0037] FIGS. **7**A-**7**F. Plot of average methylation Z score at each residue in LB (A, B and C) or WB (D, E and F) DNA from 77 female subjects of function of smoking status. The CpG residues are in order from left to right. The hatched bar indicates the residues from the first promoter island. The open bar indicates the TSS region. Group A and D. Female daily smokers (n=24). Group B and E. Females who have quit smoking (n=15). Group C and F. Females who have never smoked daily (n=38).

[0038] FIG. **8**. The sequence of the AX2R promoter associated CpG island according to the UCSC Genome Browser, Build 18. The area corresponding to the probes listed in Table 11 are highlighted and boxed. The CpG residues in the island are numbered 1 through 37 and correspond to the numbers given in Table 12.

[0039] FIG. **9**. Comparison of the smoking associated differential methylation signatures (average non-smoker beta-value minus average smoker beta value) for lymphoblast (red) and pulmonary macrophage (blue) DNA. The relative position of the 146 probes listed in Appendix A on the X-axis with the position of AHRR exons 4 (left) through 11 (right) being

noted. Please also note that exon 7 and 8 are sufficiently close to represented by a single arrow.

DETAILED DESCRIPTION

[0040] DNA Methylation

[0041] DNA does not exist as naked molecules in the cell. For example, DNA is associated with proteins called histones to form a complex substance known as chromatin. Chemical modifications of the DNA or the histones alter the structure of the chromatin without changing the nucleotide sequence of the DNA. Such modifications are described as "epigenetic" modifications of the DNA. Changes to the structure of the chromatin can have a profound influence on gene expression. If the chromatin is condensed, factors involved in gene expression may not have access to the DNA, and the genes will be switched off. Conversely, if the chromatin is "open," the genes can be switched on. Some important forms of epigenetic modification are DNA methylation and histone deacetylation. DNA methylation is a chemical modification of the DNA molecule itself and is carried out by an enzyme called DNA methyltransferase. Methylation can directly switch off gene expression by preventing transcription factors binding to promoters. A more general effect is the attraction of methyl-binding domain (MBD) proteins. These are associated with further enzymes called histone deacetylases (HDACs), which function to chemically modify histones and change chromatin structure. Chromatin-containing acetylated histones are open and accessible to transcription factors, and the genes are potentially active. Histone deacetylation causes the condensation of chromatin, making it inaccessible to transcription factors and causing the silencing of genes.

[0042] CpG islands are short stretches of DNA in which the frequency of the CpG sequence is higher than other regions. The "p" in the term CpG indicates that cysteine ("C") and guanine ("G") are connected by a phosphodiester bond. CpG islands are often located around promoters of housekeeping genes and many regulated genes. At these locations, the CG sequence is not methylated. By contrast, the CG sequences in inactive genes are usually methylated to suppress their expression.

[0043] About 56% of human genes and 47% of mouse genes are associated with CpG islands. Often, CpG islands overlap the promoter and extend about 1000 base pairs downstream into the transcription unit. Identification of potential CpG islands during sequence analysis helps to define the extreme 5' ends of genes, something that is notoriously difficult with cDNA-based approaches.

[0044] The methylation of a CpG island can be determined by the art worker using any method suitable to determine such methylation. For example, the art worker can use a bisulfite reaction-based method for determining such methylation.

[0045] The present invention provides methods to determine the nucleic acid methylation of MAOA or AHRR of a patient in order to predict the clinical course and eventual outcome of patients suspected of being predisposed or of having a substance use disorder. Previously, the only way to determine possible diagnoses was through subjective psychiatric evaluations. The present methods provide an objective component to diagnosis process.

[0046] Nicotine dependence is the physical vulnerability of a person's body to the chemical nicotine, which is potently addicting when delivered by various tobacco products. Smoke from cigarettes, cigars and pipes contains thousands of chemicals, including nicotine. Nicotine is also found in chewing tobacco. Alcohol dependence is the physical vulnerability of a person's body to the chemical ethyl alcohol.

[0047] In particular, in certain embodiments of the invention, the methods may be practiced as follows. A sample, such as a blood sample, is taken from a patient. In certain embodiments, a single cell type, e.g., lymphocytes, basophils, or monocytes isolated from the blood, may be isolated for further testing. The DNA is harvested from the sample and examined to determine if the MAOA region and/or the AHRR region is methylated. For example, the DNA of interest can be treated with bisulfite to deaminate unmethylated cytosine residues to uracil. Since uracil base pairs with adenosine, thymidines are incorporated into subsequent DNA strands in the place of unmethylated cytosine residues during subsequence PCR amplifications. Next, the target sequence is amplified by PCR, and probed with a MAOA- or AHRR-specific probe. Only DNA from the patient that was methylated will bind to the probe. A specific profile associates with a specific condition. For example, certain methylated CpG islands in MAOA are found with women having nicotine dependence (or are predisposed to having nicotine dependence), and certain methylated CpG islands in MAOA are found with women having alcohol dependence (or are predisposed to having alcohol dependence). Namely, methylated CpG islands 18, 42, 48, 52, 64-69 and 77 (in MAOA) are associated with nicotine dependence, and methylated CpG islands 27, 38, 41 and 48 (in MAOA) are associated with alcohol dependence.

[0048] Methods of determining the patient nucleic acid profile are well known to the art worker and include any of the well-known detection methods. Various PCR methods are described, for example, in *PCR Primer: A Laboratory Manual*, Dieffenbach 7 Dveksler, Eds., Cold Spring Harbor Laboratory Press, 1995. Other analysis methods include, but are not limited to, nucleic acid quantification, restriction enzyme digestion, DNA sequencing, hybridization technologies, such as Southern Blotting, etc., amplification methods such as Ligase Chain Reaction (LCR), Nucleic Acid Sequence Based Amplification (NASBA), Self-sustained Sequence Replication (SSR or 3SR), Strand Displacement Amplification (SDA), and Transcription Mediated Amplification (TMA), Quantitative PCR (qPCR), or other DNA analyses, as well as RT-PCR, in vitro translation, Northern blotting, and other RNA analyses. In another embodiment, hybridization on a microarray is used.

[0049] As used herein, the term "nucleic acid probe" or a "probe specific for" a nucleic acid means a nucleic acid sequence that has at least about 80%, e.g., at least about 90%, e.g., at least about 95% contiguous sequence identity or homology to the nucleic acid sequence encoding the targeted sequence of interest. A probe (or oligonucleotide or primer) of the invention has at least about 7-50, e.g., at least about 10-40, e.g., at least about 15-35, nucleotides. The oligonucleotide probes or primers of the invention may comprise at least about seven nucleotides at the 3' of the oligonucleotide that have at least about 80%, e.g., at least about 85%, e.g., at least about 90% contiguous identity to the targeted sequence of interest.

[0050] "Northern analysis" or "Northern blotting" is a method used to identify RNA sequences that hybridize to a known probe such as an oligonucleotide, DNA fragment, cDNA or fragment thereof, or RNA fragment. The probe is labeled with a radioisotope such as ^{32}P, by biotinylation or with an enzyme. The RNA to be analyzed can be usually

electrophoretically separated on an agarose or polyacrylamide gel, transferred to nitrocellulose, nylon, or other suitable membrane, and hybridized with the probe, using standard techniques well known in the art.

[0051] "Stringent conditions" are those that (1) employ low ionic strength and high temperature for washing, for example, 0.015 M NaCl/0.0015 M sodium citrate (SSC); 0.1% sodium lauryl sulfate (SDS) at 50° C., or (2) employ a denaturing agent such as formamide during hybridization, e.g., 50% formamide with 0.1% bovine serum albumin/0.1% Ficoll/0.1% polyvinylpyrrolidone/50 mM sodium phosphate buffer at pH 6.5 with 750 mM NaCl, 75 mM sodium citrate at 42° C. Another example is use of 50% formamide, 5×SSC (0.75 M NaCl, 0.075 M sodium citrate), 50 mM sodium phosphate (pH 6.8), 0.1% sodium pyrophosphate, 5×Denhardt's solution, sonicated salmon sperm DNA (50 μg/ml), 0.1% SDS, and 10% dextran sulfate at 42° C., with washes at 42° C. in 0.2×SSC and 0.1% SDS. Other examples of stringent conditions are well known in the art.

[0052] The term "nucleic acid" refers to deoxyribonucleotides or ribonucleotides and polymers thereof in either single- or double-stranded form, made of monomers (nucleotides) containing a sugar, phosphate and a base that is either a purine or pyrimidine. Unless specifically limited, the term encompasses nucleic acids containing known analogs of natural nucleotides that have similar binding properties as the reference nucleic acid and are metabolized in a manner similar to naturally occurring nucleotides. Unless otherwise indicated, a particular nucleic acid sequence also encompasses conservatively modified variants thereof (e.g., degenerate codon substitutions) and complementary sequences, as well as the sequence explicitly indicated. Specifically, degenerate codon substitutions may be achieved by generating sequences in which the third position of one or more selected (or all) codons is substituted with mixed-base and/or deoxyinosine residues. The terms "nucleic acid," "nucleic acid molecule," or "polynucleotide" are used interchangeably and may also be used interchangeably with gene, cDNA, DNA and/or RNA encoded by a gene.

[0053] The term "nucleotide sequence" refers to a polymer of DNA or RNA which can be single-stranded or double-stranded, optionally containing synthetic, non-natural or altered nucleotide bases capable of incorporation into DNA or RNA polymers. A DNA molecule or polynucleotide is a polymer of deoxyribonucleotides (A, G, C, and T), and an RNA molecule or polynucleotide is a polymer of ribonucleotides (A, G, C and U).

[0054] A "gene," for the purposes of the present disclosure, includes a DNA region encoding a gene product, as well as all DNA regions which regulate the production of the gene product, whether or not such regulatory sequences are adjacent to coding and/or transcribed sequences. The term "gene" is used broadly to refer to any segment of nucleic acid associated with a biological function. Genes include coding sequences and/or the regulatory sequences required for their expression. Accordingly, a gene includes, but is not necessarily limited to, promoter sequences, terminators, translational regulatory sequences such as ribosome binding sites and internal ribosome entry sites, enhancers, silencers, insulators, boundary elements, replication origins, matrix attachment sites and locus control regions. For example, "gene" refers to a nucleic acid fragment that expresses mRNA, functional RNA, or specific protein, including regulatory sequences. "Functional RNA" refers to sense RNA, antisense RNA, ribozyme RNA, siRNA, or other RNA that may not be translated but yet has an effect on at least one cellular process. "Genes" also include nonexpressed DNA segments that, for example, form recognition sequences for other proteins. "Genes" can be obtained from a variety of sources, including cloning from a source of interest or synthesizing from known or predicted sequence information, and may include sequences designed to have desired parameters.

[0055] "Gene expression" refers to the conversion of the information, contained in a gene, into a gene product. It refers to the transcription and/or translation of an endogenous gene, heterologous gene or nucleic acid segment, or a transgene in cells. In addition, expression refers to the transcription and stable accumulation of sense (mRNA) or functional RNA. Expression may also refer to the production of protein. The term "altered level of expression" refers to the level of expression in transgenic cells or organisms that differs from that of normal or untransformed cells or organisms.

[0056] A gene product can be the direct transcriptional product of a gene (e.g., mRNA, tRNA, rRNA, antisense RNA, ribozyme, structural RNA or any other type of RNA) or a protein produced by translation of an mRNA. Gene products also include RNAs which are modified, by processes such as capping, polyadenylation, methylation, and editing, and proteins modified by, for example, methylation, acetylation, phosphorylation, ubiquitination, ADP-ribosylation, myristilation, and glycosylation. The term "RNA transcript" refers to the product resulting from RNA polymerase catalyzed transcription of a DNA sequence. When the RNA transcript is a perfect complementary copy of the DNA sequence, it is referred to as the primary transcript or it may be a RNA sequence derived from posttranscriptional processing of the primary transcript and is referred to as the mature RNA. "Messenger RNA" (mRNA) refers to the RNA that is without introns and that can be translated into protein by the cell. "cDNA" refers to a single- or a double-stranded DNA that is complementary to and derived from mRNA.

[0057] A "coding sequence," or a sequence that "encodes" a selected polypeptide, is a nucleic acid molecule that is transcribed (in the case of DNA) and translated (in the case of mRNA) into a polypeptide in vivo when placed under the control of appropriate regulatory sequences. The boundaries of the coding sequence are determined by a start codon at the 5' (amino) terminus and a translation stop codon at the 3' (carboxy) terminus. A coding sequence can include, but is not limited to, cDNA from viral, prokaryotic or eukaryotic mRNA, genomic DNA sequences from viral (e.g., DNA viruses and retroviruses) or prokaryotic DNA, and especially synthetic DNA sequences. A transcription termination sequence may be located 3' to the coding sequence.

[0058] Certain embodiments of the invention encompass isolated or substantially purified nucleic acid compositions. In the context of the present invention, an "isolated" or "purified" DNA molecule or RNA molecule is a DNA molecule or RNA molecule that exists apart from its native environment and is therefore not a product of nature. An isolated DNA molecule or RNA molecule may exist in a purified form or may exist in a non-native environment such as, for example, a transgenic host cell. For example, an "isolated" or "purified" nucleic acid molecule is substantially free of other cellular material, or culture medium when produced by recombinant techniques, or substantially free of chemical precursors or other chemicals when chemically synthesized. In one embodiment, an "isolated" nucleic acid is free of

sequences that naturally flank the nucleic acid (i.e., sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived.

[0059] By "fragment" is intended a polypeptide consisting of only a part of the intact full-length polypeptide sequence and structure. The fragment can include a C-terminal deletion an N-terminal deletion, and/or an internal deletion of the native polypeptide. A fragment of a protein will generally include at least about 5-10 contiguous amino acid residues of the full-length molecule, preferably at least about 15-25 contiguous amino acid residues of the full-length molecule, and most preferably at least about 20-50 or more contiguous amino acid residues of the full-length molecule, or any integer between 5 amino acids and the full-length sequence.

[0060] Certain embodiments of the invention encompass isolated or substantially purified nucleic acid compositions. In the context of the present invention, an "isolated" or "purified" DNA molecule or RNA molecule is a DNA molecule or RNA molecule that exists apart from its native environment and is therefore not a product of nature. An isolated DNA molecule or RNA molecule may exist in a purified form or may exist in a non-native environment such as, for example, a transgenic host cell. For example, an "isolated" or "purified" nucleic acid molecule is substantially free of other cellular material or culture medium when produced by recombinant techniques, or substantially free of chemical precursors or other chemicals when chemically synthesized. In one embodiment, an "isolated" nucleic acid is free of sequences that naturally flank the nucleic acid (i.e., sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived.

[0061] "Naturally occurring" is used to describe a composition that can be found in nature as distinct from being artificially produced. For example, a nucleotide sequence present in an organism, which can be isolated from a source in nature and which has not been intentionally modified by a person in the laboratory, is naturally occurring.

[0062] "Functional RNA" refers to sense RNA, antisense RNA, ribozyme RNA, siRNA, or other RNA that may not be translated but yet has an effect on at least one cellular process.

[0063] The term "RNA transcript" refers to the product resulting from RNA polymerase catalyzed transcription of a DNA sequence. When the RNA transcript is a perfect complementary copy of the DNA sequence, it is referred to as the primary transcript or it may be a RNA sequence derived from posttranscriptional processing of the primary transcript and is referred to as the mature RNA. "Messenger RNA" (mRNA) refers to the RNA that is without introns and that can be translated into protein by the cell. "cDNA" refers to a single- or a double-stranded DNA that is complementary to and derived from mRNA.

[0064] "Regulatory sequences" and "suitable regulatory sequences" each refer to nucleotide sequences located upstream (5' non-coding sequences), within, or downstream (3' non-coding sequences) of a coding sequence, and which influence the transcription, RNA processing or stability, or translation of the associated coding sequence. Regulatory sequences include enhancers, promoters, translation leader sequences, introns, and polyadenylation signal sequences. They include natural and synthetic sequences as well as sequences that may be a combination of synthetic and natural sequences.

[0065] A "5' non-coding sequence" refers to a nucleotide sequence located 5' (upstream) to the coding sequence. It is present in the fully processed mRNA upstream of the initiation codon and may affect processing of the primary transcript to mRNA, mRNA stability or translation efficiency.

[0066] A "3' non-coding sequence" refers to nucleotide sequences located 3' (downstream) to a coding sequence and may include polyadenylation signal sequences and other sequences encoding regulatory signals capable of affecting mRNA processing or gene expression. The polyadenylation signal is usually characterized by affecting the addition of polyadenylic acid tracts to the 3' end of the mRNA precursor.

[0067] The term "translation leader sequence" refers to that DNA sequence portion of a gene between the promoter and coding sequence that is transcribed into RNA and is present in the fully processed mRNA upstream (5') of the translation start codon. The translation leader sequence may affect processing of the primary transcript to mRNA, mRNA stability or translation efficiency.

[0068] A "promoter" refers to a nucleotide sequence, usually upstream (5') to its coding sequence, which directs and/or controls the expression of the coding sequence by providing the recognition for RNA polymerase and other factors required for proper transcription. "Promoter" includes a minimal promoter that is a short DNA sequence comprised of a TATA-box and other sequences that serve to specify the site of transcription initiation, to which regulatory elements are added for control of expression. "Promoter" also refers to a nucleotide sequence that includes a minimal promoter plus regulatory elements that is capable of controlling the expression of a coding sequence or functional RNA. This type of promoter sequence consists of proximal and more distal upstream elements, the latter elements often referred to as enhancers. Accordingly, an "enhancer" is a DNA sequence that can stimulate promoter activity and may be an innate element of the promoter or a heterologous element inserted to enhance the level or tissue specificity of a promoter. It is capable of operating in both orientations (normal or flipped), and is capable of functioning even when moved either upstream or downstream from the promoter. Both enhancers and other upstream promoter elements bind sequence-specific DNA-binding proteins that mediate their effects. Promoters may be derived in their entirety from a native gene, or be composed of different elements derived from different promoters found in nature, or even be comprised of synthetic DNA segments. A promoter may also contain DNA sequences that are involved in the binding of protein factors that control the effectiveness of transcription initiation in response to physiological or developmental conditions.

[0069] "Constitutive expression" refers to expression using a constitutive promoter. "Conditional" and "regulated expression" refer to expression controlled by a regulated promoter.

[0070] "Operably-linked" refers to the association of nucleic acid sequences on a single nucleic acid fragment so that the function of one of the sequences is affected by another. For example, a regulatory DNA sequence is said to be "operably linked to" or "associated with" a DNA sequence that codes for an RNA or a polypeptide if the two sequences are situated such that the regulatory DNA sequence affects expression of the coding DNA sequence (i.e., that the coding sequence or functional RNA is under the transcriptional control of the promoter). Coding sequences can be operably-linked to regulatory sequences in sense or antisense orientation.

[0071] "Expression" refers to the transcription and/or translation of an endogenous gene, heterologous gene or

nucleic acid segment, or a transgene in cells. In addition, expression refers to the transcription and stable accumulation of sense (mRNA) or functional RNA. Expression may also refer to the production of protein.

[0072] The term "altered level of expression" refers to the level of expression in cells or organisms that differs from that of normal cells or organisms.

[0073] The following terms are used to describe the sequence relationships between two or more nucleic acids or polynucleotides: (a) "reference sequence," (b) "comparison window," (c) "sequence identity," (d) "percentage of sequence identity," and (e) "substantial identity."

[0074] (a) As used herein, "reference sequence" is a defined sequence used as a basis for sequence comparison. A reference sequence may be a subset or the entirety of a specified sequence; for example, as a segment of a full-length cDNA or gene sequence, or the complete cDNA or gene sequence.

[0075] (b) As used herein, "comparison window" makes reference to a contiguous and specified segment of a polynucleotide sequence, wherein the polynucleotide sequence in the comparison window may comprise additions or deletions (i.e., gaps) compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. Generally, the comparison window is at least 20 contiguous nucleotides in length, and optionally can be 30, 40, 50, 100, or longer. Those of skill in the art understand that to avoid a high similarity to a reference sequence due to inclusion of gaps in the polynucleotide sequence a gap penalty is typically introduced and is subtracted from the number of matches.

[0076] Methods of alignment of sequences for comparison are well-known in the art. Thus, the determination of percent identity between any two sequences can be accomplished using a mathematical algorithm. Non-limiting examples of such mathematical algorithms are the algorithm of Myers and Miller (Myers and Miller, CABIOS, 4, 11 (1988)); the local homology algorithm of Smith et al. (Smith et al., Adv. Appl. Math., 2, 482 (1981)); the homology alignment algorithm of Needleman and Wunsch (Needleman and Wunsch, JMB, 48, 443 (1970)); the search-for-similarity-method of Pearson and Lipman (Pearson and Lipman, Proc. Natl. Acad. Sci. USA, 85, 2444 (1988)); the algorithm of Karlin and Altschul (Karlin and Altschul, Proc. Natl. Acad. Sci. USA, 87, 2264 (1990)), modified as in Karlin and Altschul (Karlin and Altschul, Proc. Natl. Acad. Sci. USA 90, 5873 (1993)).

[0077] Computer implementations of these mathematical algorithms can be utilized for comparison of sequences to determine sequence identity. Such implementations include, but are not limited to: CLUSTAL in the PC/Gene program (available from Intelligenetics, Mountain View, Calif.); the ALIGN program (Version 2.0) and GAP, BESTFIT, BLAST, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Version 8 (available from Genetics Computer Group (GCG), 575 Science Drive, Madison, Wis., USA). Alignments using these programs can be performed using the default parameters. The CLUSTAL program is well described by Higgins et al. (Higgins et al., CABIOS, 5, 151 (1989)); Corpet et al. (Corpet et al., Nucl. Acids Res., 16, 10881 (1988)); Huang et al. (Huang et al., CABIOS, 8, 155 (1992)); and Pearson et al. (Pearson et al., Meth. Mol. Biol., 24, 307 (1994)). The ALIGN program is based on the algorithm of Myers and Miller, supra. The BLAST programs of Altschul et al. (Altschul et al., JMB, 215, 403 (1990)) are based on the algorithm of Karlin and Altschul supra.

[0078] Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information. This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold. These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are then extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always >0) and N (penalty score for mismatching residues; always <0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when the cumulative alignment score falls off by the quantity X from its maximum achieved value, the cumulative score goes to zero or below due to the accumulation of one or more negative-scoring residue alignments, or the end of either sequence is reached.

[0079] In addition to calculating percent sequence identity, the BLAST algorithm also performs a statistical analysis of the similarity between two sequences. One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a test nucleic acid sequence is considered similar to a reference sequence if the smallest sum probability in a comparison of the test nucleic acid sequence to the reference nucleic acid sequence is less than about 0.1, less than about 0.01, or even less than about 0.001.

[0080] To obtain gapped alignments for comparison purposes, Gapped BLAST (in BLAST 2.0) can be utilized. Alternatively, PSI-BLAST (in BLAST 2.0) can be used to perform an iterated search that detects distant relationships between molecules. When utilizing BLAST, Gapped BLAST, PSI-BLAST, the default parameters of the respective programs (e.g., BLASTN for nucleotide sequences, BLASTX for proteins) can be used. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, an expectation (E) of 10, a cutoff of 100, M=5, N=−4, and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength (W) of 3, an expectation (E) of 10, and the BLOSUM62 scoring matrix. Alignment may also be performed manually by inspection.

[0081] For purposes of the present invention, comparison of nucleotide sequences for determination of percent sequence identity to the promoter sequences disclosed herein may be made using the BlastN program (version 1.4.7 or later) with its default parameters or any equivalent program. By "equivalent program" is intended any sequence comparison program that, for any two sequences in question, generates an alignment having identical nucleotide or amino acid residue matches and an identical percent sequence identity when compared to the corresponding alignment generated by the program.

[0082] (c) As used herein, "sequence identity" or "identity" in the context of two nucleic acid or polypeptide sequences makes reference to a specified percentage of residues in the two sequences that are the same when aligned for maximum correspondence over a specified comparison window, as mea-

sured by sequence comparison algorithms or by visual inspection. When percentage of sequence identity is used in reference to proteins it is recognized that residue positions which are not identical often differ by conservative amino acid substitutions, where amino acid residues are substituted for other amino acid residues with similar chemical properties (e.g., charge or hydrophobicity) and therefore do not change the functional properties of the molecule. When sequences differ in conservative substitutions, the percent sequence identity may be adjusted upwards to correct for the conservative nature of the substitution. Sequences that differ by such conservative substitutions are said to have "sequence similarity" or "similarity." Means for making this adjustment are well known to those of skill in the art. Typically this involves scoring a conservative substitution as a partial rather than a full mismatch, thereby increasing the percentage sequence identity. Thus, for example, where an identical amino acid is given a score of 1 and a non-conservative substitution is given a score of zero, a conservative substitution is given a score between zero and 1. The scoring of conservative substitutions is calculated, e.g., as implemented in the program PC/GENE (Intelligenetics, Mountain View, Calif.).

[0083] (d) As used herein, "percentage of sequence identity" means the value determined by comparing two optimally aligned sequences over a comparison window, wherein the portion of the polynucleotide sequence in the comparison window may comprise additions or deletions (i.e., gaps) as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which the identical nucleic acid base or amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison, and multiplying the result by 100 to yield the percentage of sequence identity.

[0084] (e)(i) The term "substantial identity" of polynucleotide sequences means that a polynucleotide comprises a sequence that has at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, or 94%, or even at least 95%, 96%, 97%, 98%, or 99% sequence identity, compared to a reference sequence using one of the alignment programs described using standard parameters. One of skill in the art will recognize that these values can be appropriately adjusted to determine corresponding identity of proteins encoded by two nucleotide sequences by taking into account codon degeneracy, amino acid similarity, reading frame positioning, and the like. Substantial identity of amino acid sequences for these purposes normally means sequence identity of at least 70%, 80%, 90%, or even at least 95%.

[0085] Another indication that nucleotide sequences are substantially identical is if two molecules hybridize to each other under stringent conditions. Generally, stringent conditions are selected to be about 5° C. lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. However, stringent conditions encompass temperatures in the range of about 1° C. to about 20° C., depending upon the desired degree of stringency as otherwise qualified herein. Nucleic acids that do not hybridize to each other under stringent conditions are still substantially identical if the polypeptides they encode are substantially identical. This may occur, e.g., when a copy of a nucleic acid is created using the maximum codon degeneracy permitted by the genetic code. One indication that two nucleic acid sequences are substantially identical is when the polypeptide encoded by the first nucleic acid is immunologically cross reactive with the polypeptide encoded by the second nucleic acid.

[0086] (e)(ii) The term "substantial identity" in the context of a peptide indicates that a peptide comprises a sequence with at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, or 94%, or even 95%, 96%, 97%, 98% or 99%, sequence identity to the reference sequence over a specified comparison window. In certain embodiments, optimal alignment is conducted using the homology alignment algorithm of Needleman and Wunsch (Needleman and Wunsch, JMB, 48, 443 (1970)). An indication that two peptide sequences are substantially identical is that one peptide is immunologically reactive with antibodies raised against the second peptide. Thus, a peptide is substantially identical to a second peptide, for example, where the two peptides differ only by a conservative substitution. Thus, the invention also provides nucleic acid molecules and peptides that are substantially identical to the nucleic acid molecules and peptides presented herein.

[0087] For sequence comparison, typically one sequence acts as a reference sequence to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are input into a computer, subsequence coordinates are designated if necessary, and sequence algorithm program parameters are designated. The sequence comparison algorithm then calculates the percent sequence identity for the test sequence(s) relative to the reference sequence, based on the designated program parameters.

[0088] As noted above, another indication that two nucleic acid sequences are substantially identical is that the two molecules hybridize to each other under stringent conditions. The phrase "hybridizing specifically to" refers to the binding, duplexing, or hybridizing of a molecule only to a particular nucleotide sequence under stringent conditions when that sequence is present in a complex mixture (e.g., total cellular) DNA or RNA. "Bind(s) substantially" refers to complementary hybridization between a probe nucleic acid and a target nucleic acid and embraces minor mismatches that can be accommodated by reducing the stringency of the hybridization media to achieve the desired detection of the target nucleic acid sequence.

[0089] "Stringent hybridization conditions" and "stringent hybridization wash conditions" in the context of nucleic acid hybridization experiments such as Southern and Northern hybridizations are sequence dependent, and are different under different environmental parameters. Longer sequences hybridize specifically at higher temperatures. The thermal melting point (Tm) is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. Specificity is typically the function of post-hybridization washes, the critical factors being the ionic strength and temperature of the final wash solution. For DNA-DNA hybrids, the T_m can be approximated from the equation of Meinkoth and Wahl (1984); T_m 81.5° C.+16.6 (log M)+0.41 (% GC)−0.61 (% form)−500/L; where M is the molarity of monovalent cations, % GC is the percentage of guanosine and cytosine nucleotides in the DNA, % form is the percentage of formamide in the hybridization solution, and L is the length of the hybrid in base pairs.

T_m is reduced by about 1° C. for each 1% of mismatching; thus, T_m, hybridization, and/or wash conditions can be adjusted to hybridize to sequences of the desired identity. For example, if sequences with >90% identity are sought, the T_m can be decreased 10° C. Generally, stringent conditions are selected to be about 5° C. lower than the T_m for the specific sequence and its complement at a defined ionic strength and pH. However, severely stringent conditions can utilize a hybridization and/or wash at 1, 2, 3, or 4° C. lower than the T_m; moderately stringent conditions can utilize a hybridization and/or wash at 6, 7, 8, 9, or 10° C. lower than the T_m; low stringency conditions can utilize a hybridization and/or wash at 11, 12, 13, 14, 15, or 20° C. lower than the T_m. Using the equation, hybridization and wash compositions, and desired temperature, those of ordinary skill will understand that variations in the stringency of hybridization and/or wash solutions are inherently described. If the desired degree of mismatching results in a temperature of less than 45° C. (aqueous solution) or 32° C. (formamide solution), the SSC concentration is increased so that a higher temperature can be used. Generally, highly stringent hybridization and wash conditions are selected to be about 5° C. lower than the T_m for the specific sequence at a defined ionic strength and pH.

[0090] An example of highly stringent wash conditions is 0.15 M NaCl at 72° C. for about 15 minutes. An example of stringent wash conditions is a 0.2×SSC wash at 65° C. for 15 minutes. Often, a high stringency wash is preceded by a low stringency wash to remove background probe signal. An example medium stringency wash for a duplex of, e.g., more than 100 nucleotides, is 1×SSC at 45° C. for 15 minutes. For short nucleotide sequences (e.g., about 10 to 50 nucleotides), stringent conditions typically involve salt concentrations of less than about 1.5 M, less than about 0.01 to 1.0 M, Na ion concentration (or other salts) at pH 7.0 to 8.3, and the temperature is typically at least about 30° C. and at least about 60° C. for long probes (e.g., >50 nucleotides). Stringent conditions may also be achieved with the addition of destabilizing agents such as formamide. In general, a signal to noise ratio of 2× (or higher) than that observed for an unrelated probe in the particular hybridization assay indicates detection of a specific hybridization. Nucleic acids that do not hybridize to each other under stringent conditions are still substantially identical if the proteins that they encode are substantially identical. This occurs, e.g., when a copy of a nucleic acid is created using the maximum codon degeneracy permitted by the genetic code.

[0091] Very stringent conditions are selected to be equal to the T_m for a particular probe. An example of stringent conditions for hybridization of complementary nucleic acids that have more than 100 complementary residues on a filter in a Southern or Northern blot is 50% formamide, e.g., hybridization in 50% formamide, 1 M NaCl, 1% SDS at 37° C., and a wash in 0.1×SSC at 60 to 65° C. Exemplary low stringency conditions include hybridization with a buffer solution of 30 to 35% formamide, 1 M NaCl, 1% SDS (sodium dodecyl sulphate) at 37° C., and a wash in 1× to 2×SSC (20×SSC=3.0 M NaCl/0.3 M trisodium citrate) at 50 to 55° C. Exemplary moderate stringency conditions include hybridization in 40 to 45% formamide, 1.0 M NaCl, 1% SDS at 37° C., and a wash in 0.5× to 1×SSC at 55 to 60° C.

[0092] In a further embodiment of the invention, there are provided articles of manufacture and kits containing probes, oligonucleotides or antibodies which can be used, for instance, for the diagnostic applications described above. The article of manufacture comprises a container with a label. Suitable containers include, for example, bottles, vials, and test tubes. The containers may be formed from a variety of materials such as glass or plastic. The container holds a composition which includes an agent that is effective for diagnostic applications, such as described above. The label on the container indicates that the composition is used for a specific diagnostic application. The kit of the invention will typically comprise the container described above and one or more other containers comprising materials desirable from a commercial and user standpoint, including buffers, diluents, filters and package inserts with instructions for use.

[0093] The probes of the present invention can be labeled using techniques known to those of skill in the art. For example, the labels used in the assays of invention can be primary labels (where the label comprises an element that is detected directly) or secondary labels (where the detected label binds to a primary label, e.g., as is common in immunological labeling). An introduction to labels (also called "tags"), tagging or labeling procedures, and detection of labels is found in Polak and Van Noorden (1997) Introduction to Immunocytochemistry, second edition, Springer Verlag, N.Y. and in Haugland (1996) Handbook of Fluorescent Probes and Research Chemicals, a combined handbook and catalogue Published by Molecular Probes, Inc., Eugene, Oreg. Primary and secondary labels can include undetected elements as well as detected elements. Useful primary and secondary labels in the present invention can include spectral labels such as fluorescent dyes (e.g., fluorescein and derivatives such as fluorescein isothiocyanate (FITC) and Oregon Green™, rhodamine and derivatives (e.g., Texas red, tetramethylrhodamine isothiocyanate (TRITC), etc.), digoxigenin, biotin, phycoerythrin, AMCA, CyDyes™, and the like), radiolabels (e.g., ^{3}H, ^{125}I, ^{35}S, ^{14}C, ^{32}P, ^{33}P), enzymes (e.g., horse-radish peroxidase, alkaline phosphatase) spectral colorimetric labels such as colloidal gold or colored glass or plastic (e.g., polystyrene, polypropylene, latex) beads. The label may be coupled directly or indirectly to a component of the detection assay (e.g., the labeled nucleic acid) according to methods well known in the art. As indicated above, a wide variety of labels may be used, with the choice of label depending on sensitivity required, ease of conjugation with the compound, stability requirements, available instrumentation, and disposal provisions. In general, a detector that monitors a probe-substrate nucleic acid hybridization is adapted to the particular label that is used. Typical detectors include spectrophotometers, phototubes and photodiodes, microscopes, scintillation counters, cameras, film and the like, as well as combinations thereof. Examples of suitable detectors are widely available from a variety of commercial sources known to persons of skill. Commonly, an optical image of a substrate comprising bound labeled nucleic acids is digitized for subsequent computer analysis.

[0094] Preferred labels include those that use (1) chemiluminescence (using Horseradish Peroxidase and/or Alkaline Phosphatase with substrates that produce photons as breakdown products) with kits being available, e.g., from Molecular Probes, Amersham, Boehringer-Mannheim, and Life Technologies/Gibco BRL; (2) color production (using both Horseradish Peroxidase and/or Alkaline Phosphatase with substrates that produce a colored precipitate) (kits available from Life Technologies/Gibco BRL, and Boehringer-Mannheim); (3) hemifluorescence using, e.g., Alkaline Phosphatase and the substrate AttoPhos (Amersham) or other sub-

strates that produce fluorescent products, (4) Fluorescence (e.g., using Cy-5 (Amersham), fluorescein, and other fluorescent labels); (5) radioactivity using kinase enzymes or other end-labeling approaches, nick translation, random priming, or PCR to incorporate radioactive molecules into the labeled nucleic acid. Other methods for labeling and detection will be readily apparent to one skilled in the art.

[0095] Fluorescent labels are highly preferred labels, having the advantage of requiring fewer precautions in handling, and being amendable to high-throughput visualization techniques (optical analysis including digitization of the image for analysis in an integrated system comprising a computer). Preferred labels are typically characterized by one or more of the following: high sensitivity, high stability, low background, low environmental sensitivity and high specificity in labeling. Fluorescent moieties, which are incorporated into the labels of the invention, are generally are known, including Texas red, dixogenin, biotin, 1- and 2-aminonaphthalene, p,p'-diaminostilbenes, pyrenes, quaternary phenanthridine salts, 9-aminoacridines, p,p'-diaminobenzophenone imines, anthracenes, oxacarbocyanine, merocyanine, 3-aminoequilenin, perylene, bis-benzoxazole, bis-p-oxazolyl benzene, 1,2-benzophenazin, retinol, bis-3-aminopyridinium salts, hellebrigenin, tetracycline, sterophenol, benzimidazolylphenylamine, 2-oxo-3-chromen, indole, xanthen, 7-hydroxycoumarin, phenoxazine, calicylate, strophanthidin, porphyrins, triarylmethanes, flavin and many others. Many fluorescent labels are commercially available from the SIGMA Chemical Company (Saint Louis, Mo.), Molecular Probes, R&D systems (Minneapolis, Minn.), Pharmacia LKB Biotechnology (Piscataway, N.J.), CLONTECH Laboratories, Inc. (Palo Alto, Calif.), Chem Genes Corp., Aldrich Chemical Company (Milwaukee, Wis.), Glen Research, Inc., GIBCO BRL Life Technologies, Inc. (Gaithersberg, Md.), Fluka ChemicaBiochemika Analytika (Fluka Chemie AG, Buchs, Switzerland), and Applied Biosystems™ (Foster City, Calif.), as well as many other commercial sources known to one of skill.

[0096] Means of detecting and quantifying labels are well known to those of skill in the art. Thus, for example, where the label is a radioactive label, means for detection include a scintillation counter or photographic film as in autoradiography. Where the label is optically detectable, typical detectors include microscopes, cameras, phototubes and photodiodes and many other detection systems that are widely available.

[0097] The present invention is further detailed in the following Examples, which are offered by way of illustration and are not intended to limit the invention in any manner. Standard techniques well known in the art or the techniques specifically described below are utilized. All patent and literature references cited in the present specification are hereby incorporated by reference in their entirety.

Example 1

MAOA Methylation is Associated with Nicotine and Alcohol Dependence

[0098] Over the past several years, it has become increasingly evident that gene-environment interactions (GxE) and residual gene-environment correlations (rGE) have a prominent role in the etiology of most common behavioral illnesses. However, the exact processes underlying these interactions and the extent of their relative contributions are unclear. At the molecular level, epigenetic phenomena such as DNA methylation and histone modification are thought to contribute to these processes. Unfortunately, empirical data to support this hypothesis at behaviorally relevant loci have been scarce.

[0099] Two candidate loci at which epigenetic phenomena may participate in GxE, rGE or E effects are the Serotonin Transporter (SLC6A4) and Monoamine Oxidase A (MAOA). The protein products of both of these two loci play prominent roles in regulating serotonergic and monoaminergic transmission, respectively. These moderating roles have come under increasing scrutiny due to recent studies which have demonstrated prominent GxE effects for depression at SLC6A4 (Caspi et al., *Science.* 301(5631), 386-9 (2003)) and for aggression at MAOA (Kim-Cohen et al., *Mol Psychiatry.* 11(10), 903-913 (2006); Caspi et al., *Science.* 297(5582), 851-4 (2002)). Hence, there is a great deal of curiosity as to the mechanisms through which E or GxE effects could influence biological processes at these loci.

[0100] One mechanism through which GxE or E effects could become manifest at the molecular level is through altering relevant gene expression through methylation of gene promoters in response to environmental stressors. In the initial study of the relationship between promoter methylation and behavioral phenomena, the inventors conducted quantitative methylation analyses of the SLC6A4 associated promoter CpG island and demonstrated that methylation of this promoter is both sex dependent and associated with increased vulnerability to major depression. However, whether there is a similar promoter associated CpG island at MAOA, and if it exists, whether its methylation has behavioral consequences was unclear.

[0101] Two types of disorders that could potentially be influenced by methylation induced changes in MAOA activity are Antisocial Personality Disorder (ASPD) and substance use disorders (SUD). Already, genetic variation in a variable nucleotide repeat (VNTR) located immediately upstream of the MAOA minimal central promoter has been associated with different vulnerability to ASPD and two forms of SUD: alcohol dependence (AD) and nicotine dependence (ND).

[0102] In this report, using a set of similar techniques to prior methylation and gene expression analyses of SCL6A4 (Philibert et al., *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, (2007)) and the resources of the Iowa Adoption Studies (IAS), a large longitudinal adoption study focusing on the role of GxE effects in SUD, the inventors examined the relationship of MAOA genotype and methylation to SUD and ASPD.

Methods

[0103] The procedures used in the IAS have been described in detail elsewhere (Yates et al., *Drug and Alcohol Dependence.* 41(1), 9 (1996)). Briefly, the IAS is a case and control adoption study of G, E and GxE effects in SUD and ASPD. This study, founded by Remi Cadoret, contrasts the outcomes of 475 adoptees from the State of Iowa who are at high biological risk for SUD or ASPD (i.e., one of their biological parents was severely affected) with those of 475 adoptees who were not at biological risk for either SUD or ASPD. After birth, each of these adoptees was randomly placed in an adoptive home. Since their inception in the study, the adoptees and their adoptive environments have been serially assessed. The subjects included in this pilot study were the first 95 males and 96 females to participate in this wave of the study. The overall study design and all procedures described in this communication were approved by the University of Iowa Institutional Review Board.

[0104] Briefly, the behavioral and biological material used in these studies was obtained from subjects who participated

in the last two waves of the Iowa Adoptions Studies (IAS). In both of these waves, each subject was interviewed with a version of the Semi Structured Assessment for the Genetics of Alcoholism, Version 2 (SSAGA-II) (Bucholz et al., *J Stud Alcohol.* 55(2), 149-58 (1994)). In addition, in the latest round of the study, phlebotomy was performed on each of the participants. Symptom counts and categorical diagnoses for each of the disorders (ASPD, AD, ND) were derived from SSAGA-II data using the individual dependence or personality disorder criteria from DSM-IV (Association, AP, *Diagnostic and Statistical Manual of Mental Disorder, Fourth Edition.* 1994, Washington D.C.: American Psychiatric Association), with the highest total symptom count from these two interviews being defined as the lifetime symptom count.

[0105] RNA and DNA used in the studies were derived from lymphoblast cell lines using biomaterial contributed by the participants. These lymphoblast cell lines were prepared using standard EBV transfection techniques from the specimens contributed by the study participants (Klaus, G G B, *Lymphocytes: A practical Approach.* 1987, Oxford: IRL Press. 149-162). Total RNA was prepared from lymphoblast using a Midi RNA purification kit from Invitrogen™ (Carlsbad, Calif.) according to the manufacturer's instructions. DNA was prepared from lymphoblast cell pellets using cold protein precipitation (Lahiri et al., *Nucleic Acids Research.* 19(19), 5444 (1991)).

[0106] PCR amplification of the MAOA variable nucleotide repeat (VNTR) polymorphism was conducted using the method of Sabol and colleagues (Sabol et al., *Hum Genet.* 103(3), 273-9 (1998)). The resulting PCR products were electrophoresed on a 6% non-denaturing polyacrylamide gel and imaged using silver staining (Merril et al., *Analytical Biochemistry.* 156(1), 96-110 (1986)). The resulting alleles were compared to internal standards and the genotypes were called by two individuals blind to affected status.

[0107] RTPCR was conducted as previously described (Philibert et al., *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, (2007); Bradley et al., *Am J Med Genet B Neuropsychiatr Genet.* 136(1), 58-61 (2005)). Briefly, RNA was reverse transcribed using an Applied Biosystems™ cDNA archiving kit (Foster City, Calif.). Then, 12.5 ng aliquots of cDNA were robotically dispensed and RTPCR performed using reagents from Applied Biosystems™ including primer-probe sets for MAOA (Hs 00165140) and the endogenous control loci GAPDH (from the GAPDH Control kit) and LDHA (Hs 00855332).

[0108] The existence, location, size and sequence of the MAOA CpG islands were determined using the default browser settings of the University of California Genome Browser (UCSC) website (world-wide-web at genome.ucsc.edu). The sequences for these islands are freely available from the website or from the authors on request.

[0109] Quantitative methylation analyses for each of the samples at these CpG residues were conducted by Sequenom® Inc. (San Diego, Calif.) as previously described (Philibert et al., [erratum appears in Mol Psychiatry 1999 March; 4(2):197.]. *Molecular Psychiatry.* 3(4), 303-9 (1998)). First, aliquots of purified DNA were treated using bisulfite modification (Frommer et al., *Proceedings of the National Academy of Sciences.* 89(5), 1827-1831 (1992)). Treatment of DNA with bisulfite deaminates unmethylated cytosine residues to uracil. Since uracil base pairs with adenosine, thymidines are incorporated into subsequent DNA strands in the place of unmethylated cytosine residues during subsequent PCR amplifications. Next, contigs covering the CpG islands (see FIG. **1**) were PCR amplified. Because of the size of the region, the CpG enriched regions were PCR amplified in four separate reactions. The primers for each of those PCR amplifications are as follows: Amplicon A (from BP 43398925 to 43399181): F-TTA AAG AAT GAA AGT ATT AGG TTG AGA GTT (SEQ ID NO:1) and R-ATA CCC ACT CTT AAA AAC CAA CCC C (SEQ ID NO:2); Amplicon B (from BP 43399430 to 43399858): F-GGG TGT TGA ATT TTG AGG AGA AG (SEQ ID NO:3) and R-AAA ACA CAA CTA CCC AAA TCC C (SEQ ID NO:4); Amplicon C (from BP 43400453 to 43400805): F-GGG GAG TTG ATA GAA GGG TTT TTT TTA T (SEQ ID NO:5) and R-TAT ATC TAC CTC CCC CAA TCA CAC C (SEQ ID NO:6) and Amplicon D (from BP 43400486 to 43400035): F-AAA GGG TGG GAA GGA TTT TTT TAT TAA TT (SEQ ID NO:7) and R-CAT CCT CAA TAT CCA ACT TCC CCT A (SEQ ID NO:8) using standard touchdown PCR conditions (Philibert et al., *Am J Med Genet B Neuropsychiatr Genet.* 144(1), 101-5 (2007)). Methylation ratios for each of the CpG residues (methyl CpG/total CpG) were then determined using a MassARRAY™ mass spectrometer using proprietary peak picking and spectra interpretation tools (Ehrich et al., *Proc Natl Acad Sci USA.* 102(44), 15785-90 (2005); Ehrich et al., *Nucleic Acids Res.* 35(5), e29 (2007)).

[0110] The data were analyzed using the JMP (version 7; SAS Institute, Cary, S.C.) using Pearson's correlation coefficients, regression [analysis of variance (ANOVA) and ordinal logistic regression (OLR)] or Chi-square testing as indicated in the text (Fleiss, J L, *Statistical Methods for Rates and Proportions.* 2nd ed. 1981, New York, N.Y.: John Wiley & Sons Inc.). All tests were two-tailed and all analyses were conducted by gender.

Results

[0111] The characteristics of the IAS subjects who contributed the biomaterials to this study are given in Table 1. In total, 96 female and 95 male subjects provided biomaterials for the study. The male subjects were significantly older than the female subjects (t-test, $p<0.002$) and had a significantly higher symptom count for ASPD (Chi-Square, $p<0.001$).

TABLE 1

Demographic and Clinical Characteristics of the IAS Subjects

	Male	Female
N	95	96
Age (years ± SD)	42.4 ± 8.5	38.8 ± 6.8
Ethnicity		
White	87	91
African American	5	2
White of Hispanic Origin	2	1
Other	1	2

DSM IV Symptom Counts

#	ASPD		AD		ND	
Symptoms	M	F	M	F	M	F
0	18	41	35	49	47	50
1	26	30	25	25	4	6
2	21	9	16	13	10	7
3	7	7	11	3	15	8
4	10	5	2	2	6	14

TABLE 1-continued

Demographic and Clinical Characteristics of the IAS Subjects

5	9	3	3	3	8	8
6	4	3	2	0	3	3
7	0	0	1	0	2	0

[0112] The MAOA VNTR genotypes for the subjects are given in Table 2. The testing for Hardy Weinberg equilibrium in the female subjects was unremarkable.

TABLE 2

MAOA VNTR Genotype.

Genotype	Female Subjects	Male Subjects*
2, 2	0	1
2, 4	1	—
3, 3	18	34
3, 4	41	—
3, 5	1	—
3.5, 3.5	0	1
3.5, 4	1	—
4, 4	31	59
4, 5	3	—

*Male subjects are hemizygous with respect to this X-chromosome locus.

[0113] Sequence analysis of MAOA demonstrated the presence of two CpG islands in the gene (FIG. **1**). The first island, stretching from by 43398975 to by 43399158, contains 18 CpG residues and is approximately 1200 bp upstream of the transcription start site for MAOA. The second CpG island begins at by 43399493 and contains 70 CpG residues. Exon 1 of MAOA is wholly contained within the CpG island with the transcription start site (TSS) for the gene occurring between CpG residues 64 and 65. The MAOA VNTR is found between the two CpG islands.

[0114] The average methylation ratio at each of these residues is shown in FIG. **2**. As the figure demonstrates, females have consistently higher methylation ratios at each CpG residue than males (who are hemizygous for this gene). Please note that secondary to methodological limitations with respect to the ability of the mass spectrograph to resolve individual residues, the values for CpG residues, 1-2, 5-7, 11-12, 19-20, 30-31, 43-44, 55-57, 67-68, 72-73, and 79-80 are shown as aggregates.

[0115] The interrelationships of MAOA methylation between individual residues for each gender were studied. The correlation between methylation was higher between residues in the smaller 5' CpG island than it was in between residues in the larger CpG island that encompasses Exon 1. Of particular potential interest, methylation of the two residues immediately flanking the TSS, CpG 64 and 65, is poorly correlated with methylation throughout the rest of the island. However, methylation at the residues CpG 58-63 and CpG 66-70 is highly inter-correlated.

[0116] In order to test the hypothesis that MAOA genotype influences the amount of methylation, the relationship of average methylation to genotype at the VNTR (FIG. **3**) for each gender was analyzed. There was a trend for female 3,3 homozygotes to have a higher average methylation than female 4,4 homozygotes (43.3%±3.8 vs 40.9%±5.2; $p<0.10$). There was no significant difference between males hemizygous for the 3 repeat allele as compared to those with the 4 allele although the arithmetic difference was in the same direction (9.0±3.7 vs 8.3±2.6; $p<0.32$).

[0117] The relationship between symptom counts for ASPD, AD and ND with average methylation for each gender was then analyzed using ordinal regression analysis. There was no relationship between ASPD and overall methylation for neither men (OLR, $p<0.37$) or women (OLR, $p>0.70$). There also were not any significant relationships between average methylation and AD (OLR, $p<0.23$) and ND (OLR, $p<0.68$) in male subjects. However, there were strong relationships between average overall methylation and symptom counts for AD (OLR, $p<0.008$) and ND (OLR, $p<0.002$) in female subjects.

[0118] In order to identify the residues driving the strong correlations between overall methylation and symptom counts for AD and ND in women, the relationship between methylation at individual CpG residues and symptom counts was analyzed. With respect to former, methylation at CpG residues 27, 38, 41, and 48 were nominally significantly associated ($p<0.05$ before correction for multiple comparisons) with AD symptom count in female subjects. With respect to the latter, methylation at CpG residues 18, 42, 48, 52, 64, 65, 67-68, 69, and 77 were nominally associated ($p<0.05$ before correction for multiple comparisons) with ND symptom counts.

[0119] Finally, in an attempt to discern whether gene expression was correlated with MAOA genotype or methylation, the inventors attempted to measure MAOA gene expression using our previously described techniques (Philibert et al., *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, (2007); Philibert, et al., *Am J Med Genet B Neuropsychiatr Genet.* 144(1), 101-5 (2007); Philibert et al., *Am J Med Genet B Neuropsychiatr Genet.* 144(5), 683-90 (2007)). Unfortunately, despite several attempts, we could not reliably detect MAOA gene expression.

Discussion

[0120] In summary, it was discovered that MAOA methylation is associated with ND and AD in women, but not men. In addition, a significant relationship between ASPD and CpG methylation was not found in men or women. Finally, there was a trend for MAOA genotype to be associated with methylation in women.

[0121] The results with respect to ND are perhaps the most compelling. Review of the animal model literature shows that MAOA knockout mice exhibit impaired nicotine preference but have normal responses to other novel stimuli (Agatsuma et al., *Hum. Mol. Genet.* 15(18), 2721-2731 (2006)). Furthermore, treatment of rats with the monoamine oxidase inhibitor phenelzine enhances the discriminant stimulus effect of nicotine (Wooters et al., *Behav Pharmacol.* 18(7), 601-8 (2007)) and increases nicotine self administration (Villegier et al, *Neuropharmacology.* 52(6), 1415-25 (2007); Guillem et al., *Eur J Neurosci.* 24(12), 3532-40 (2006)). Review of the literature with respect to humans reveals that the targeting of neurotransmitter systems regulated by MAOA, using agents such as reboxetine, a selective norepinephrine reuptake inhibitor, or bupropion, which targets the dopaminergic system, have been shown to be clinically effective in the treatment of ND (Miller et al., *J Pharmacol Exp Ther.* 302(2), 687-695 (2002); George, T and Weinberger, A: Monoamine Oxidase Inhibition for Tobacco Pharmacotherapy. *Mol Ther*, (2007); David et al., *Tobacco Research.* 9(8), 821-833

(2007)). Finally, platelet MAOA activity is reduced in smokers (Berlin et al., *Int J Neuropsychopharmacol.* 4(1), 33-42 (2001)).

[0122] This evidence is made even more compelling by closer inspection and consideration of the MAOA methylation data with respect to ND in the female subjects. The control of transcription initiation is one of the major mechanisms through which cells regulate gene expression (Levine et al., *Nature.* 424(6945), 147-151 (2003)). Hence, the TSS is a frequent target of epigenetic modifications including methylation and histone modification (Kawaji et al., *Genome Biology.* 7(12), R118 (2006); Liang et al., *Proc Natl Acad Sci USA.* 101(19), 7357-62 (2004)). Therefore, it is expected that any significant changes in ND association MAOA methylation preferentially affects the MAOA TSS. This is indeed what is observed with a strong clustering of CpG residues that are either nominally significantly associated or with a trend for association ($p<0.10$) surrounding the TSS.

[0123] It is important to note that the primary outcome measure with respect to methylation in this study was overall methylation, not individual CpG residue methylation. This is because it was not known prior to the study which CpG residues might be most important.

[0124] Surprisingly, the inventors did not find any relationship between MAOA methylation and ASPD. Nor did the inventors find a significant relationship between genotype and methylation. Once again, this may simply be a function of low power. At the same time, these findings do not preclude specific GxE effects on ASPD at this locus because they did not examine the relationship of environmental factors hypothesized to elicit such effects, such as maltreatment, in this study.

[0125] In summary, it was discovered that methylation of the MAOA promoter is associated with ND and AD in females.

Example 2

The Effect of Smoking on MAOA Promoter Methylation in DNA Prepared from Lymphoblasts and Whole Blood

[0126] Monoamine Oxidase A (MAOA) plays a key role in modulating monoaminergic neurotransmission through its catabolism of dopamine, norepinephrine, epinephrine, serotonin and related neurotransmitter catabolism byproducts. The MAOA gene is located on Xp11 and consists of 15 exons that are transcribed to 4.1 kb mRNA and translated into a 527 amino acid protein (Chen Z Y et al. 1991. Structure of the human gene for monoamine oxidase type A. *Nucleic Acids Res* 19(16):4537-41). Two regulatory motifs for the gene have been previously described. The first is a 44 bp variable nucleotide repeat (VNTR) that is found approximately 1200 bp upstream of the transcription start site (TSS) (Hotamisligil G S, Breakefield X O. 1991. Human monoamine oxidase A gene determines levels of enzyme activity. *Am J Hum Genet* 49(2):383-92). The second is a set of two promoter associated CpG islands that flank either side of the VNTR (See Example 1 above).

[0127] As discussed above, it was demonstrated that increased lifetime symptom counts for Alcohol (AD) and Nicotine Dependence (ND) were associated with decreased MAOA methylation, with the effects being most prominent in women in the region of the gene surrounding the TSS. Furthermore, evidence was provided that the three-repeat (3R) allele of the VNTR was associated with increased methylation at this locus. In the present study several further questions were raise. First, were these findings simply a type I error due to multiple tests across CpG island loci? Second, given the direct pharmacological effects of nicotine consumption, is decreased methylation associated with current smoking only, or is there an effect of history of smoking as well? Third, are some regions of the promoter more important in characterizing this process? Finally, do lymphoblasts provide better or worse resolution than alternative media, such as whole blood, for the examination of epigenetic effects in substance use research?

[0128] These are important concerns because MAOA is hypothesized to play a key role in ND and other complex behavioral illnesses. MAOA inhibitors are used in the treatment of ND as well as other frequently co-morbid syndromes, such as major depression (MD) (George T P, Weinberger A H. 2008. Monoamine oxidase inhibition for tobacco pharmacotherapy. *Clin Pharmacol Ther* 83(4):619-21). Furthermore, MAOA VNTR gene-environment (GxE) interaction specific to the 3-epeat allele (3R) may be important in the etiology of antisocial conduct (Caspi A et al. 2002. Role of genotype in the cycle of violence in maltreated children. Science 297 (5582):851-4; Frazzetto G et al., 2007. Early Trauma and Increased Risk for Physical Aggression during Adulthood: The Moderating Role of MAOA Genotype. *PLoS ONE* 2(5): e486). Finally, the present researchers have recently confirmed earlier findings that a similar GxE effect specific to the 4R allele may moderate vulnerability to MD (Beach S R et al., in submission. Child Maltreatment and MAOA Genotype in Depression and Antisocial Personality Disorder: Genetic Moderation of Family Environment). Therefore, the development of a detailed understanding of the molecular underpinnings of genetic and epigenetic effects at this locus is beneficial to the understanding and treatment of complex behavioral illness.

[0129] To help accomplish this goal and more finely hone our understanding of genetic and epigenetic effects at this locus, the inventors recently re-examined the original findings using the insights derived from our prior study and the resources provided by 289 additional participants in the IAS.

Methods

[0130] The study design and clinical measures in the IAS have been described in detail elsewhere (Yates W R et al. 1996. An adoption study of DSM-IIIR alcohol and drug dependence severity. Drug and Alcohol Dependence 41(1): 9). The behavioral and demographic data were obtained from subjects participating in the last two waves of the IAS (1997-2003; 2004-2009). In each wave, subjects were interviewed with a version of the Semi-Structured Assessment for the Genetics of Alcoholism, version 2 (SSAGA-II) (Bucholz K K et al. 1994. A new, semi-structured psychiatric interview for use in genetic linkage studies: a report on the reliability of the SSAGA. *J Stud Alcohol* 55(2):149-58). In addition, in the last wave subjects were phlebotomized to provide biomaterial for the preparation of DNA and lymphoblast cell lines. All these procedures were approved by the University of Iowa Institutional Review Board.

[0131] The clinical and laboratory methods used in this study are very similar to those used previously. With respect to the behavioral data, symptom counts and categorical diagnoses for nicotine dependence were derived from SSAGA-II data using criteria from DSM-IV (American Psychiatric

Association 1994). The highest total symptom count from these two interviews was defined as the lifetime symptom count. Smoking status was also determined using SSAGA data. Those who denied a history of daily smoking at both interviews were classified as "non-smokers." Those subjects who were daily smokers at the time of the first interview, but had totally quit at time 2 were classified as "quitters." Those who smoked daily at the time of both interviews were classified as "continuous smokers."

[0132] DNA from two different cellular sources was used in this study. The lymphoblast (LB) DNA for all 289 subjects was prepared from cell lines using blood contributed by the participants. These cell lines were derived using standard EBV transfection techniques (Klaus G G B. 1987. Lymphocytes: A practical Approach. Oxford: IRL Press. p. 149-162) and the DNA was harvested using the method of Lahiri and Schnabel (Lahiri D K, Schnabel B. 1993. DNA isolation by a rapid method from human blood samples: effects of MgCl2, EDTA, storage time, and temperature on DNA yield and quality. *Biochem Genet* 31(7-8):321-8). For a subset of the female subjects (n=78), we also analyzed DNA that was prepared from the whole blood sample (WB DNA) drawn at the same time as the specimen used to prepare the cell lines. This DNA was also extracted using the method of Lahiri and Schnabel (Lahiri D K, Schnabel B. 1993. DNA isolation by a rapid method from human blood samples: effects of $MgCl_2$, EDTA, storage time, and temperature on DNA yield and quality. *Biochem Genet* 31(7-8):321-8), and the methylation signatures of both types of DNA were determined at the same time.

[0133] Genotyping of the MAOA variable nucleotide repeat (VNTR) polymorphism was conducted as previously described (see Example 1 above). Quantitative methylation determination was performed under contract by Sequenom® Inc. (San Diego, Calif.) using the same methods previously described (Philibert R A et al. 1998. Association of an X-chromosome dodecamer insertional variant allele with mental retardation. [erratum appears in *Mol Psychiatry* 1999 March; 4(2):197.]. *Molecular Psychiatry* 3(4):303-9). First, aliquots of purified DNA underwent bisulfite modification. The modified DNA samples were then used as a template for the PCR amplification of three contigs covering the MAOA promoter islands using standard touchdown conditions (Philibert R et al. 2007. Serotonin transporter mRNA levels are associated with the methylation of an upstream CpG island. *Am J Med Genet B Neuropsychiatr Genet* 144(1):101-5). Amplicon A stretches from BP 43398925 to 43399181, covers CpG residues 1 to 18, and uses the following primers: F-TAA AGA ATG AAA GTA TTA GGT TGA GAG TT (SEQ ID NO:1) and R-ATA CCC ACT CTT AAA AAC CAA CCC C (SEQ ID NO:2). Amplicon B stretches from BP 43399430 to 43399858, covers CpG residues 19 to 45, and uses the following primers: F-GGG TGT TGA ATT TTG AGG AGA AG (SEQ ID NO:3) and R-AAA CAC AAC TAC CCA AAT CCC (SEQ ID NO:4); Amplicon C stretches from BP 43400453 to 43400805, covers CpG residues 46 to 74, and uses the following primers: F-GGG GAG TTG ATA GAA GGG TTT TTT TTA T (SEQ ID NO:5) and R-TAT ATC TAC CTC CCC CAA TCA CAC C (SEQ ID NO:6). A fourth contig that covered CpG 75-88 was not used in this study because the residues in this amplicon were neither correlated with methylation in other amplicons nor with substance use in Example 1.

[0134] After amplification, the methylation ratios for each of the CpG residues (methyl CpG/total CpG) in these contigs were then determined using a MassARRAY™ system mass spectrometer (Sequenom®). These data were then analyzed with proprietary peak picking and spectra interpretation tools to generate the methyl CpG/total CpG ratios (Ehrich M et al. 2005. Quantitative high-throughput analysis of DNA methylation patterns by base-specific cleavage and mass spectrometry. *Proc Natl Acad Sci USA* 102(44):15785-90; Ehrich M, et al. 2007. A new method for accurate assessment of DNA quality after bisulfite treatment. *Nucleic Acids Res* 35(5): e29). The peak for some residues could not be de-convoluted by the spectral interpretation tools. In those cases (CpG 5-7, 8-9, 11-12, 19-20, 30-31, 61-62, 67-68, 72-73), the value for each residue is presented as an average of the aggregated values. In addition, no signal could be reliably observed for CpG residues 24, 26, and 28.

[0135] Because the methylation data had differing means and standard deviations at each loci, all methylation data were Z-transformed before comparison to genotype or clinical data. All data were analyzed using the JMP (version 7; SAS Institute, Cary, S.C.) using Pearson's correlation coefficients, regression, analysis of variance (ANOVA), T-tests, and ordinal logistic regression (OLR)] as indicated in the text (Fleiss 1981). Factor analyses were conducted using SAS Version 9.1 (SAS Institute, Cary, N.C.). For analyses of VNTR genotype data, genotypes that contained uncommon alleles (i.e. 2, 3.5, and 5 repeats) were excluded and the remaining genotype data were analyzed using an additive model. All tests were two-tailed.

Results

[0136] The basic behavioral and demographic characteristics of this cohort of 289 IAS subjects are given in Table 3. As with the prior cohort, most of the subjects are White and well into adulthood. The male subjects do not differ from the female subjects with respect to age nor ethnicity. Consistent with the study design of the IAS, the sample is enriched for behavioral illness with 100 subjects reporting 3 or more lifetime criteria for ND.

TABLE 3

Demographic and Clinical Characteristics of the IAS Subjects

	Male	Female
N	125	164
Age (years ± SD)	41.1 ± 7.7	40.9 ± 7.7
Ethnicity		
White	117	155
African American	3	2
White of Hispanic Origin	4	4
Other	1	3

DSM IV Symptom Counts for ND

#Symptoms	Males	Females
0	55	84
1	9	12
2	8	11
3	12	9
4	20	18
5	9	19
6	11	9
7	1	2

[0137] The genotype distribution of the subjects is given in Table 4. No relationship emerged between the MAOA VNTR genotype and lifetime symptom count for ND for males ($p<0.98$, OLR) or females ($p<0.19$, OLR).

TABLE 4

MAOA VNTR Genotype.

Genotype	Female Subjects (n = 164)	Male Subjects (n = 125)
2, 2	0	1
2, 4	0	—
3, 3	21	43
3, 3.5	1	—
3, 4	64	—
3, 5	1	—
3.5, 3.5	0	4
3.5, 4	2	—
4, 4	71	76
4, 5	2	—
5, 5	0	1
Unknown	2	—

*male subjects are hemizygous with respect to this X-chromosome locus.

[0138] The untransformed sex averaged methyl CpG/total CpG ratio for each residue is given in FIG. **4**. The first CpG island contains 18 CpG residues and begins approximately ~1200 bp before the transcription start site of MAOA. The VNTR lies between the two CpG islands. The second island consists of 70 CpG residues, the first 56 residues of which were measured in this study. The TSS is located between CpG residues 64 and 65. Overall, males have a average methylation ratio (methyl CpG/total CpG) of 7.2% and females have an average methylation ratio of 34.8%.

[0139] Not surprisingly, because MAOA is an X chromosome gene, females consistently had a higher average methylation ratio at every CpG residue. Ethnicity was not associated with average methylation. However, a trend emerged for increasing age to be associated with increasing methylation in females ($p<0.07$; ANOVA) but not males ($p<0.30$; ANOVA).

[0140] Average and Locus Specific Methylation. The relationship between the Z-transformed average methylation ratios across the 74 residues examined and VNTR genotype is shown in FIG. **5**. The average Z-transformed methylation ratio was greater in DNA from heterozygous females (3R,4R) than in DNA from 4R homozygotes ($p<0.04$; T-test). Although the directionality of differential methylation was consistent with prior findings, hemizygous males and homozygous females for the 3R allele did not have significantly higher average amounts of methylation than did their 4R counterparts ($p<0.24$ and $p<0.20$, respectively).

[0141] Next, the inventors examined the relationship between global or TSS region specific methylation, which was defined as being the average of Z-transformed values for residues CpG 61-70, and lifetime ND symptom count for all 289 subjects. Although the pattern of relationships was similar to prior findings, the relationship between global methylation and lifetime ND symptom count was not statistically significant for males ($p<0.19$) or females ($p<0.12$). However, before correction for multiple comparisons, eight individual CpG residues (CpG 22, 25, 32, 36, 39, 64, 65 and 69), including three in the TSS region, were nominally associated ($p<0.05$) with ND symptom for the male subjects but no such relationships emerged in the female subjects.

[0142] Because the inventors noted that a substantial number of subjects had quit smoking, yet were still counted as affected using the lifetime symptom count criterion, the inventors next examined current smoking status for 274 subjects whose smoking status could be easily classified. First, for these analyses of current smoking status, those who denied a history of smoking one or more days per week were designated non-smokers (male n=59 and female n=83). "Daily smokers" were defined as those who smoked 7 days per week at the times of both the first and second interviews (male n=42 and female n=45). Finally, "quitters" were defined as those subjects who smoked daily at the time of the first interview, but denied smoking regular smoking (1 or more days per week) at the time of the second interview (male n=20 and female n=27). The 15 subjects excluded from these three groups were removed because either they were never truly daily smokers at both interviews (i.e., did not smoke every day; n=10), did not fully quit smoking (n=4), or started smoking after the first interview (n=1).

[0143] Using these definitions of smoking status, the examined the relationship between global and site-specific methylation and current daily smoking status. The distribution of the differential methylation at each residue for male and female "lifetime daily smokers," "quitters" and non-smokers" is illustrated in FIG. **6**. The results are most marked for the male subjects. As compared to non-smokers, smokers had lower amounts of methylation globally ($p<0.02$; T-test) and at the transcription start site ($p<0.009$; T-test) with 7 residues meeting nominal significance level before correction for multiple comparisons. As FIG. **6** demonstrates, smoking is associated with a pervasive decrease in methylation across the second larger CpG island with particular consistency in two areas. The first is from CpG 19 to CpG 32. The second is from CpG 55 to CpG 69, a region that includes the TSS. In contrast, the methylation pattern in those male subjects who quit in the five years prior to the blood draw is decidedly mixed across both islands, with both elevated and decreased methylation at particular residues. Finally, in those male subjects without a history of daily smoking, the net methylation is pervasively increased across the larger CpG island, but somewhat mixed and perhaps decreased overall in the first CpG island.

[0144] The methylation pattern in LB DNA from female smokers is similar to that of the male smokers but less intense and consistent. A clear contrast is seen between the amount and pattern methylation observed in those females who quit smoking as compared to those who never smoked, with a trend for reduced overall methylation ($p<0.08$; T-test) and a significant reduction of methylation at the TSS ($p<0.04$; T-test) in those who quit.

[0145] Factor Analytic Results. To determine whether methylation data could aggregated in a meaningful way, the inventors used the FACTOR procedure in the SAS computer program (SAS Institute, Cary N.C.) to factor analyze the set of CpG residues for which >95% of both male and female participants had scores. This approach provided a stable three dimensional factor structure accounting for 39% of the reliable variance. The inventors used a varimax rotation to identify regions of covariation in degree of methylation. Use of the three factor scores has the advantage of summarizing the reliable signal in the data, while minimizing the number of separate contrasts required to describe effects, which enhances the signal to noise ratio in the data.

[0146] The three regions identified by the factor analysis were: Factor 1 (CpG 19-CpG 45), Factor 2 (CpG 58-CpG 74), and Factor 3 (CpG 1-CpG 18). Use of average scores across

the identified region provided a similar pattern of results as use of factor scores. Therefore, factor scores were used in all analyses reported below.

[0147] Replicating and extending the analyses reported above for genotype, the inventors found that methylation was greater for heterozygous (3R,4R) or homozygous (4R) females, but the effect was confined to Factor 3 (i.e., CpG 1-CpG 18), $F(1,137)=4.50$, $p<0.05$. The average factor scores for the three groups across CpG 1-18 were (−0.17 vs. 0.23 vs. −0.10) for homozygous 4R, heterozygous 3R,4R and homozygous 3R females respectively. The inventors also found a significant effect of genotype for males, but in this case the effect was confined to factor 1 (CpG 19-CpG 45), $F(1,122)=5.25$, $p<0.03$. The average factor scores for the two groups across CpG 19-45 were (−0.11 vs. 0.20) for the hemizygous 4R vs. 3R males respectively. For both males and females, the 4R allele was associated with significantly less methylation.

[0148] Replicating and extending the analysis of global methylation effects, a significant association between methylation in the region of CpG 19-45 and days smoking at time 1 ($p<0.002$) and time 2 ($p<0.02$) for males was found. A significant association also emerged between days smoking and methylation in the region around the TSS (i.e., Factor 2; CpG 56-74) for males, but only at time 1 ($p<0.02$). For females, the only significant association emerged between factor 3 (CpG 1-18) and smoking at time 1 ($p<0.04$). For ND symptom count, we found trends for males $p<0.07$, for factor 1 (CpG 19-45) and $p<0.1$ for factor 2 (CpG 56-74), but no significant associations for females.

[0149] The inventors next replicated and extended the analyses contrasting continuous smokers, quitters, and non-smokers. The inventors found significant group differences for males in methylation of factor 1 (CpG 19-45), $F(2, 117)=5.46$, $p<0.01$, and factor 2 (CpG 56-74), $F(2, 117)=3.91$, $p<0.05$. The average factor scores for the three groups across CpG 19-45 were (−0.19 vs. −0.25 vs. −0.15) for non-smokers, continuous smokers, and quitters respectively. Males who never smoked had the highest level of methylation whereas continuous smokers had the least, and quitters were intermediate. For Factor 2 (CpG 56-74), those who never smoked also had the highest methylation, but the quitters had the least. The average factors scores for the three groups were (0.15 vs. −0.15 vs. −0.29) for non-smokers, continuous smokers, and quitters, respectively. For females, only Factor 3 (CpG 1-18) reliably differentiated the groups $F(2,150)=3.04$, $p=0.05$. Females who never smoked had the highest methylation and those who had quit had the lowest. The average factors scores for the three groups were (0.15 vs. −0.15 vs. −0.41) for non-smokers, continuous smokers, and quitters, respectively.

[0150] Comparison of Lymphoblasts to Whole Blood. Finally, because there is considerable controversy in the field as to which source(s) of DNA can or should be used in methylation studies, the inventors next compared the relationship of smoking status to ND in 78 of the female subjects included in the above analyses using DNA prepared from whole blood (WB) or from the lymphoblast line (LB) derived from the same sample of blood. Each set of samples had a similar amount of overall methylation (LB 33.3% vs WB 34.0%, $p<0.45$; T-test). The distribution with respect to VNTR allele status was virtually identical (data not shown). With respect to smoking status, there was a trend for decreased overall methylation in DNA of smokers (n=24) as compared to that from non-smoking females (n=38) when the DNA was derived from the lymphoblasts ($p<0.09$; T-test). However, there was no difference when the same comparison was performed using DNA prepared from whole blood ($p<0.89$; T-test). To gain a better understanding of this, the inventors plotted the methylation signatures at each residue for those who were daily smokers, recently quit, or who had never smoked. Although the same patterns are present in the DNA from both sources, visual inspection of the methylation plots demonstrates greater consistency and intensity of the differential methylation patterns in the DNA derived from lymphoblasts as compared to that from whole blood.

[0151] To compare the results of the methylation results from the two sources of DNA in a more quantitative manner, the inventors next examined average methylation in the three regions identified in the factor analysis using a 3 (smoking status) by 2 (LB vs. WB DNA) ANOVA for each region. As before, only the region identified by Factor 3 (CpG 1-18) reliably differentiated the three smoking status groups $F(2, 74)=4.61$, $p<0.02$. There was no interaction with type of assessment (WB or LB DNA) for this region, suggesting that, given enough observations, a method using either source of DNA would have identified the pattern—even though the spread of the distribution of means was slightly more pronounced for LB than for WB samples (0.21, −0.09, −0.38 vs. 0.15, −0.03, −0.34 for non-smokers, continuous smokers, and quitters, respectively for LB vs. WB samples). There was, however, a trend toward significance for the interaction of smoking status with assessment method for Factor 1 (CpG 19-45) $F(2,74)=2.45$, $p<0.1$, suggesting that the two approaches might lead to somewhat different conclusions for that region of the CpG island. In particular, the pattern of means for the three smoking status groups was (0.10, −0.11, −0.03 vs. −0.14, 0.16, 0.11) for non-smokers, continuous smokers, and quitters respectively for LB vs. WB samples, indicating a reversal of the relative positions of never smokers and quitters in average level of methylation in this region depending on which assessment method was used.

Discussion

[0152] In summary, using another sample of subjects from the IAS, the inventors replicated and extended their previous findings to show that a significant portion of the methylation signature status at MAOA is associated with current smoking status, that quitting has an effect on methylation status, and that gender and region of the CpG island examined are also important for accurate specification of associations between smoking and level of methylation. The inventors also examined an important methodological issue by using methylation data on the same subjects using two different sources of DNA, and by examining relationships using a factor analytic approach to reduce the number of dimensions required to describe the methylation results.

[0153] The current data provide compelling evidence that the methylation status of the two CpG islands associated with the MAOA promoter is dependent upon smoking status. The real question is why? The answer may be to increase the amount of MAOA protein that is produced. Previous work by others has shown that acute exposure to smoke decreased human brain MAOA activity (Fowler J S, et al. 1996. Brain monoamine oxidase A inhibition in cigarette smokers. *Proc Natl Acad Sci USA* 93(24):14065-9), and that this decrease in protein activity may be a direct pharmacological/toxicological effect of substances in tobacco smoke (Berlin I, Anthenelli R M. 2001. Monoamine oxidases and tobacco smoking *Int J*

Neuropsychopharmacol 4(1):33-42; Fowler J S et al. 2003. Monoamine Oxidase and Cigarette Smoking *NeuroToxicology* 24(1):75-82.). Since promoter methylation, particularly at the TSS generally decreases mRNA transcription, it seems plausible that the association of decreased methylation with increasing ND symptom count could result from the attempt of the cell to upregulate MAOA RNA production in the face of increased MAOA protein turnover or inhibition caused by smoking.

[0154] Whereas this appears to readily explain the contrast in methylation between current smokers and non-smokers, this rationale does not fully explain the effect of "quitting" on MAOA methylation that does not appear to lead to uniform changes and a return to methylation levels similar to those who never smoked. Indeed, on most indices the quitters were as different from non-smokers as the continuous smokers, albeit more variable in their methylation profiles. However, at this time one should be cautious in the interpretation of this portion of these findings. The window of time for "quitting" for these subjects used in this study was rather large and it is highly likely that the subjects differed significantly between one another with respect to total time of smoking abstinence. Therefore, aggregating all "quitters" together in analyses may be insensitive important heterogeneity in this group. Still, taken at face value, these data suggest that the process of returning to non-smoking methylation status may be a lengthy one and that the process may be dynamic at the molecular level as well as at the clinical level.

[0155] The finding that female 4R homozygotes have significantly lower methylation than 3R,4R heterozygotes and arithmetically lower methylation than 3R homozygotes is consistent with the inventors' prior work in which they showed a trend for the 4R homozygotes to have lower average methylation than 3R homozygotes (40.9% vs 43.3%; $p<0.10$). In unpublished data from that analysis, the average methylation of the 3,4 heterozygotes was only slightly less than that of the 3R homozygotes (42.9%). Hence, when the data is pooled, it is clear that the average methylation of the 4R homozygotes is significantly lower than that of both 3,4 heterozygotes as well as the 3R homozygotes. In addition, this pattern was found for males when factor scores were examined, albeit only for CpG residues in the region from 19-45. Unfortunately, at this time, there is not a good explanation for the observation that the "low activity" 3R allele is associated with greater average methylation overall, and in particular, the region of the first CpG island. The inventors' expectation going into these studies was that the 4R allele would have greater methylation than the 3R allele in order to compensate for the greater amount of gene transcription that has been shown in most, but not all, transfection studies (Beach S R et al., in submission, Child Maltreatment and MAOA Genotype in Depression and Antisocial Personality Disorder: Genetic Moderation of Family Environment; Cirulli E T, Goldstein D B. 2007. In vitro assays fail to predict in vivo effects of regulatory polymorphisms. *Hum Mol Genet.* 16(16):1931-1939; Guo G et al. 2008. The VNTR 2 repeat in MAOA and delinquent behavior in adolescence and young adulthood: associations and MAOA promoter activity. *Eur J Hum Genet* 16(5):626-34; Sabol S Z et al. 1998. A functional polymorphism in the monoamine oxidase A gene promoter. *Hum Genet* 103(3):273-9). But this is not the case, suggesting that more complex regulatory processes may be at work or that transfections of these MAOA alleles does not fully capture the transcriptional complexity present at this locus.

[0156] Lymphoblast cultures are homogenous cell lines that are derived from long lived peripheral β-lymphocyte populations and are relatively unaffected by acute changes in the health status of the host (Hao Z, Rajewsky K. 2001. Homeostasis of peripheral B cells in the absence of B cell influx from the bone marrow. *J Exp Med* **194**(8):1151-64; Tough D F, Sprent J. 1995. Lifespan of lymphocytes. *Immunol Res* 14(1):1-12). Others have demonstrated that the epigenetic signature is preserved in lymphoblasts (Monks S A et al. 2004. Genetic inheritance of gene expression in human cell lines. *Am J Hum Genet* 75(6):1094-105; Morello F et al. 2004. Differential Gene Expression of Blood-Derived Cell Lines in Familial Combined Hyperlipidemia. *Arterioscler Thromb Vasc Biol* 24(11):2149-2154). The present observation of nearly identical amounts of total methylation and allele specific methylation in the WB and LB samples further supports this supposition. In contrast, there are several reasons to believe that the methylation signatures in WB DNA may be more variable. Peripheral white blood cells are a varying mixture of neutrophils, lymphocytes, eosinophils, basophils and monocytes, each of which probably has a slightly different methylation signature. The composition of this cell mix can change suddenly. In particular, the neutrophil portion of this mixture is subject to marked swings in population secondary to margination of these cells to the blood stream in response to processes such as stress, infection or drug ingestion (e.g., lithium). Because these processes are associated with changes in neutrophil protein and gene expression signatures (Bussiere F I et al. 2002. Stress protein expression cDNA array study supports activation of neutrophils during acute magnesium deficiency in rats. *Magnes Res* 15(1-2):37-42; Macdonald J, Galley H F, Webster N R. 2003. Oxidative stress and gene expression in sepsis. *Br J Anaesth* 90(2):221-232), it is likely that as part of these processes, changes in methylation signatures also occur, leading to greater variability in WB than LB DNA. In light of this source of variability in the constituent elements of WB DNA and the likelihood that the various cell types in blood differ slightly in their methylation signatures, it is reasonable to assume that WB DNA may have greater variability in its methylation signature. However, this does not mean it should not be used in these types of studies. Careful review of FIG. **7** demonstrates that the same patterns are evident in both sources of DNA and the current data are from just one locus.

[0157] The apparent differences in the methylation profiles with respect to smoking status are intriguing. Although the inventors initially analyzed only overall and TSS specific methylation, one advantage of using factor analytic scores is that they provide a potentially useful way of defining and then summarizing methylation for all regions of the CpG island, allowing better specification of possible differences between groups and between genders. For example, average factor analytic scores for males show an orderly transition from decreased to increased methylation as a function of smoking status that is most apparent for Factor 1 comprising the region from CpG 19 to CpG 45. For females, non-smokers also demonstrate the highest methylation, but this is most evident on Factor 3 comprising the region from CpG 1 to CpG 18. Both male and female quitters demonstrated lower levels of methylation than did non-smokers on Factor 2 (i.e., the region containing the TSS) with continuous smokers being intermediate (−0.29 vs −0.14 vs 0.14 for males; −0.26, −0.02, 0.13 for females), suggesting that effects of smoking status at the TSS

may be more similar than different for males and females, and that quitting smoking may be associated with lowered methylation for both.

Example 3

Genome-Wide Methylation Analysis

[0158] Genome wide methylation analyses were conducted using lymphoblast DNA from 10 well controls, 8 subjects with active alcohol dependence, 7 subjects with active Nicotine Dependence and 4 subjects with active Cannabis Dependence from the Iowa Adoption Studies. Briefly, 10 μg of highly purified lymphoblast DNA from each subject was digested with to completion with MseI, purified and a 300 ng aliquot (input) stored for further analysis. Then, 5 μg aliquots of each sample were denatured at 95° C. for 10 min, and subsequently rapidly chilled. The denatured DNA was then resuspended in immunoprecipitation buffer, then sequentially immunoprecipitated with mouse anti 5-methylcytosine (Abcam, USA), and sheep anti-mouse IgG antibodies. The resulting immunoprecipitated DNA was then cleaved from the precipitated complex by overnight proteinase K digestion and purified. Then, aliquots of both the input and enriched (immunoprecipitated) DNA were amplified with a Whole Gene Amplification-2 (WGA-2) kit (Sigma, USA) according to manufacturer's instructions. The resulting DNA was purified and quantified. Then, 5 μg aliquots of resulting amplified DNA samples were shipped to Roche-Nimblegen (Indianapolis) for labeling and hybridization under contract. In short, the input and enriched DNA samples were labeled with Cy-3 and Cy-5, respectively and then matching specimens were be hybridized to the 385 K NimbleGen promoter array and scanned.

[0159] Analysis of Genome wide data: Cy3-Cy5 ratios for probe were computed, $\log_2$ transformed, then scaled by subtracting the bi-weight mean from each value for each feature. The resulting values were then analyzed in relation to all features and directly neighboring features by fixed window Kolmogorov-Smirnov test to identify significantly differentially regulated promoter regions in subjects. The resulting peak scores for each differentially region for each subject were exported and the results from cases and controls contrasted using standard t-tests to determine differentially regulated individual gene promoter regions in type of substance use syndrome.

[0160] Three tables are given (Table 5, 6 and 7) with respect to the identity of gene promoter region differentially regulated in Nicotine, Alcohol and Cannabis Dependence. These promoter-associated islands are listed according to their HUGO identification of the gene to which they are associated.

TABLE 5

Genes whose methylation is differentially regulated in DNA from subjects with active Nicotine Dependence as compared to DNA from well Controls.

ACCN3	KIRREL2	PAX3
ATP6V0A4	KLK9	PDCD4
C10orf39	LCA10	PNCK
C10orf53	LIMS3	PPAN
C21orf123	LOC155006	PRAME
CACNA1G	LOC285095	PSG7
CCDC49	LOC643274	PTPRT
CCNC	LOC645811	RIBC1
CD8A	LOC646836	RP11-159H20.4

TABLE 5-continued

Genes whose methylation is differentially regulated in DNA from subjects with active Nicotine Dependence as compared to DNA from well Controls.

CDH16	LOC653176	S100A1
CIDEB	LOC653700	S100A13
CLEC10A	LPHN1	SCN5A
CYP2B6	LTB4R	SELS
DOK2	LW-1	SH3PX3
DYDC1	MAFG	SLC35E1
EFNA3	MAGEA4	SPIN-2
FABP6	MATK	TAC3
FAM107B	MCART1	TBX4
FKBPL	MGC4728	TCOF1
FLJ32569	MYADML	TNFSF9
FLJ40365	N/A	TOMM40
FLJ43870	NCAM1	VMD2L1
GALP	NCR3	WFIKKN1
GJB1	NOXO1	ZCCHC13
GMPPA	NUDT1	ZFP64
HOXA5	NUMB	ZNF274
HRASLS2	OPN1LW	ZNF320
KCNQ1DN	OR2B11	ZNF516
KHSRP	OR6V1	
KIAA1843	PAQR5	

TABLE 6

Genes whose methylation is differentially regulated in DNA from subjects with active Alcohol Dependence as compared to DNA from well Controls.

ZGP1	FBXW5	MLX	SHARPIN
BHLHB8	FLJ40448	MPG	SLITRK4
C6orf26	FRG1	N/A	SNAPC2
C20orf70	GIYD2	OPRS1	SULT1A3
CACNA1S	KIAA1875	PANX2	TBX2
CMTM2	KLK8	PIP5KL1	THTPA
COL6A2	LOC339047	PPP1CA	TMEM101
COLEC11	LOC440354	PRSS27	TMEM121
CSAG2	LOC642628	PSG3	TRIM17
ELF3	LOC644122	REEP6	TYRO3
FAM3A	LOC645598	RFNG	

TABLE 7

Genes whose methylation is differentially regulated in DNA from subjects with active Cannabis Dependence as compared to DNA from well Controls.

IL32	ZNF42	SPANXA1	TOMM40
PEO1	FNDC8	TMEM88	SERTAD3
LOC653210	FAM84A	C14orf120	IER2
C7orf21	PTPN20A	IGFBP6	ARHGEF1
PTPN20A	RBP5	ACSS2	PEO1
KRT17	LOC642358	FXYD1	BMF
PTPN20B	PAQR8	CMTM1	KIAA0310
LOC653107	DNAI1	H2AFB3	DNAJC19
GIYD2	LOC653680	KIAA0892	SEPT6
FTL	HSPA1A	ZNF409	MAGED4
FLJ21767	SIRT2	IRF7	HSD11B1L
LOC653107	UBOX5	LOC653257	GIYD2
CSH2	TUBA2	RTN2	EGLN2
CSAG3A	KCNK7	KCNK7	PRCP
MGC12760	ZNF580	SULT1A3	CYBA
MUC4	RNASEH2A	N/A	TUSC4
RRP22	SCT	LOC339123	BCKDK
S100A13	LOC653210	LOC644083	GH2
TRIM74	LOC644733	ATP6V0C	BOLA2
BAD	RBM10	PAIP2	PITRM1
CSAG3A	DARC	LOC653483	LOC401019

TABLE 7-continued

Genes whose methylation is differentially regulated in DNA from subjects with active Cannabis Dependence as compared to DNA from well Controls.

ANKRD25	LIME1	MEIS3	FAM39A
SNRPN	MAGEA2B	MAGED4	CKAP1
FKRP	FLJ21767	CA5BL	GMFG
RNF126	SNRPN	ARRB2	CYP2D6
ITGB4BP	COX6B2	RIPK3	BAG1
PEO1	C1orf142	CKS1B	Rgr
LRDD	C7orf21	TSEN34	LOC653107
CHMP5	CACNA1C	SFTPC	DEDD2
CRAT	CSAG3A	OBP2B	PITX3
FAM39A	MGC12760	BAGE	
FAM39A	FLJ36046	CRYAB	
ECM1	PEO1	LOC389833	

Example 4

Methylation Profiling of Nicotine Dependence

[0161] Nicotine dependence (ND) is one of the largest public health challenges in the developed world. Despite extensive treatment and prevention efforts, approximately 20% of U.S. adults still smoke on a daily basis which results in 440,000 premature deaths and $92 billion dollars of economic costs annually (Center for Disease Control. 2005 Annual Smoking-Attributable Mortality, Years of Potential Life Lost, and Productivity Losses—United States, 1997-2001. Morbidity and Mortality Weekly 54(25):625-628; Center for Disease Control. 2009. State-Specific Prevalence and Trends in Adult Cigarette Smoking—United States, 1998-2007. JAMA 302(3):250-252.) Not surprisingly, a large number of studies have been conducted to identify the genetic and environmental factors associated with smoking. While the analyses of both types of factors have been informative and useful in the provision of better treatment and prevention measures, the rate of smoking in the general population may have reached a nadir and in fact may be increasing in young adults (Kumra V, Markoff B A. 2000. WHO'S SMOKING NOW?: The Epidemiology of Tobacco Use in the United States and Abroad. Clinics in Chest Medicine 21(1):1-9.). Hence, there is increased urgency to understand the biology underlying ND. Unfortunately, even though recent genome wide analyses have clearly identified significant genetic variation for ND (Bierut L J, Madden P A, Breslau N, Johnson E O, Hatsukami D, Pomerleau O F, Swan G E, Rutter J, Bertelsen S, Fox L and others. 2007. Novel genes identified in a high-density genome wide association study for nicotine dependence. Hum Mol Genet 16(1):24-35; Vink J M, Smit A B, de Geus E J, Sullivan P, Willemsen G, Hottenga J J, Smit J H, Hoogendijk W J, Zitman F G, Peltonen L and others. 2009. Genome-wide association study of smoking initiation and current smoking Am J Hum Genet 84(3):367-79.), the majority of the biological vulnerability for initiation and maintenance of smoking behaviors remains unexplained.

[0162] Recently there has been an increasing appreciation that a portion of the biology responsible for the initiation and maintenance of smoking behaviors may be epigenetic. Over the past two years, a number of studies have demonstrated that smoking itself induces biological changes at loci such as monoamine oxidase A (MAOA) and monoamine oxidase B (MAOB) which are known to be important in human behavior (Fowler J S, Logan J, Wang G-J, Volkow N D. 2003. Monoamine Oxidase and Cigarette Smoking NeuroToxicology 24(1):75-82; Fowler J S, Volkow N D, Wang G J, Pappas N, Logan J, Shea C, Alexoff D, MacGregor R R, Schlyer D J, Zezulkova I and others. 1996b. Brain monoamine oxidase A inhibition in cigarette smokers. Proc Natl Acad Sci USA 93(24):14065-9). Whereas some of the biological effects are known to be due to the direct effects of cigarette smoke (Yu P H, Boulton A A. 1987. Irreversible inhibition of monoamine oxidase by some components of cigarette smoke. Life Sci 41(6):675-82), it is also becoming evident that smoking may directly affect the methylation status of genes (Breton C V, Byun H-M, Wenten M, Pan F, Yang A, Gilliland F D. 2009. Prenatal Tobacco Smoke Exposure Affects Global and Gene-Specific DNA Methylation. Am J Respir Crit Care Med: 200901-0135OC; Philibert R, Beach S R, Gunter T, Brody G H, Madan A. 2009. The Effect of Smoking on MAOA Promoter Methylation in DNA Prepared from Lymphoblasts and Whole Blood. Am J Med Genet B Neuropsychiatr Genet September 23; [Epub ahead of print]). These findings are intriguing because altered DNA methylation is an integral part of the biological processes in the carcinogenic pathway (Tessema M, Yu Y Y, Stidley C A, Machida E O, Schuebel K E, Baylin S B, Belinsky S A. 2009. Concomitant promoter methylation of multiple genes in lung adenocarcinomas from current, former and never smokers. Carcinogenesis 30(7): 1132-1138) and they suggest the possibility that methylation may also affect behaviorally relevant genes.

[0163] There is strong support for the hypothesis that smoking alters DNA methylation of behaviorally relevant genes at the Xp13 locus containing MAOA and MAOB. Monoamine oxidase activity is essential for the normal catabolism of monoaminergic neurotransmitters. Classically, disruption of this oxidase activity is associated with aberrant behavior, especially aggression (Brunner H G, Nelen M, Breakefield X O, Ropers H H, van Oost B A. 1993. Abnormal behavior associated with a point mutation in the structural gene for monoamine oxidase A. Science 262(5133):578-80). Since that seminal discovery, there has been an increasing body of evidence, including a set of elegant neuroimaging analyses by Volkow and associates, that implicates altered MAOA and MAOB protein activity in the CNS and non-CNS pathophysiology associated with smoking (Alia-Klein N, Goldstein R Z, Kriplani A, Logan J, Tomasi D, Williams B, Telang F, Shumay E, Biegon A, Craig I W and others. 2008. Brain Monoamine Oxidase A Activity Predicts Trait Aggression. J Neurosci 28(19):5099-5104; Fowler J S, Logan J, Wang G-J, Volkow N D. 2003. Monoamine Oxidase and Cigarette Smoking NeuroToxicology 24(1):75-82; Fowler J S, Volkow N D, Wang G J, Pappas N, Logan J, MacGregor R, Alexoff D, Shea C, Schlyer D, Wolf A P and others. 1996a. Inhibition of monoamine oxidase B in the brains of smokers. Nature 379(6567):733-6; Fowler J S, Volkow N D, Wang G J, Pappas N, Logan J, Shea C, Alexoff D, MacGregor R R, Schlyer D J, Zezulkova I and others. 1996b. Brain monoamine oxidase A inhibition in cigarette smokers. Proc Natl Acad Sci USA 93(24):14065-9). Some of these changes in smoking associated monoamine oxidase activity are secondary to direct effects of smoke (Yu P H, Boulton A A. 1987. Irreversible inhibition of monoamine oxidase by some components of cigarette smoke. Life Sci 41(6):675-82). However, an emerging literature has indicated that altered epigenetic regulation of both of these genes may also be playing a role in altering monoamine oxidase activity (Launay J-M, Del Pino M, Chironi G, Callebert J, Peoc'h K, Megnien J-L, Mallet J, Simon

A, Rendu F. 2009. Smoking Induces Long-Lasting Effects through a Monoamine-Oxidase Epigenetic Regulation. PLoS ONE 4(11):e7959; Philibert R, Beach S R, Gunter T, Brody G H, Madan A. 2009. The Effect of Smoking on MAOA Promoter Methylation in DNA Prepared from Lymphoblasts and Whole Blood. Am J Med Genet B Neuropsychiatr Genet September 23; [Epub ahead of print]; Philibert R A, Gunter T D, Beach S R, Brody G H, Madan A. 2008. MAOA methylation is associated with nicotine and alcohol dependence in women. Am J Med Genet B Neuropsychiatr Genet 147B(5): 565-70). Taken together with a recent genome wide study of methylation of the effects of maternal prenatal smoking (Breton C V, Byun H-M, Wenten M, Pan F, Yang A, Gilliland F D. 2009. Prenatal Tobacco Smoke Exposure Affects Global and Gene-Specific DNA Methylation. Am J Respir Crit Care Med: 200901-0135OC) and studies by others indicating that altered methylation loci such as OPRM1, DAT and SNCA (Bonsch D, Lenz B, Kornhuber J, Bleich S. 2005. DNA hypermethylation of the alpha synuclein promoter in patients with alcoholism. Neuroreport 16(2):167-70; Hillemacher T, Frieling H, Hartl T, Wilhelm J, Kornhuber J, Bleich S. 2009. Promoter specific methylation of the dopamine transporter gene is altered in alcohol dependence and associated with craving. Journal of Psychiatric Research 43(4):388-392; Nielsen D A, Yuferov V, Hamon S, Jackson C, Ho A, Ott J, Kreek M J. 2008. Increased OPRM1 DNA Methylation in Lymphocytes of Methadone-Maintained Former Heroin Addicts. Neuropsychopharmacology) in other addictive behaviors, a nascent literature is emerging that supports the assertion that various addictive substances may alter DNA methylation at a broad number of loci relevant to behavior, and that better understanding changes in methylation may enhance our understanding of the biology of addiction.

[0164] Since DNA methylation is a major mechanism through which gene expression and ultimately behavior is regulated, these findings also suggest that smoking induced altered DNA methylation may be in part responsible for some of the processes which maintain smoking as well as some of the other behavioral phenomena associated with smoking, such as increased risk for panic disorder (Isensee B, Wittchen H U, Stein M B, Hofler M, Lieb R. 2003. Smoking increases the risk of panic: findings from a prospective community study. Arch Gen Psychiatry 60(7):692-700). Capturing a broader understanding of that biology may generate critical insights that may be important to the development of better treatment and prevention measures for smoking and associated phenomena. Therefore, in order to begin the facilitation of this better understanding of this altered DNA methylation on a more systematic basis, an analysis was conducted of DNA methylation at 18,028 promoter associated CpG islands using lymphoblast DNA from 23 actively smoking ND subjects and 18 age and ethnicity matched controls from the Iowa Adoption Studies.

Methods

[0165] The design and diagnostic measures used in the IAS have been extensively described previously and all have been approved by the University of Iowa Institutional Review Board (Yates W R, Cadoret R J, Troughton E, Stewart M A. 1996. An adoption study of DSM-IIIR alcohol and drug dependence severity. Drug and Alcohol Dependence 41(1): 9). The clinical data used in the study was derived from the latest two rounds of structured interviews conducted in our studies (1999-2003 and 2004-2009). The core instrument for these studies was an adaptation of the Structured Assessment for the Genetic Studies of Alcoholism, version 2 (SSAGA-II) (Bucholz K K, Cadoret R, Cloninger C R, Dinwiddie S H, Hesselbrock V M, Nurnberger J I, Jr., Reich T, Schmidt I, Schuckit M A. 1994. A new, semi-structured psychiatric interview for use in genetic linkage studies: a report on the reliability of the SSAGA. J Stud Alcohol 55(2):149-58). The lifetime symptom counts for ND and Fagerstrom Tests for Nicotine Dependence (FTND) scores were compiled from this data using DSM-IV criteria and published scales as previously described (Heatherton T F, Kozlowski L T, Frecker R C, Fagerstrom K O. 1991. The Fagerstrom Test for Nicotine Dependence: a revision of the Fagerstrom Tolerance Questionnaire. Br J Addict 86(9):1119-27; Philibert R A, Ryu G Y, Yoon J G, Sandhu H, Hollenbeck N, Gunter T, Barkhurst A, Adams W, Madan A. 2007. Transcriptional profiling of subjects from the Iowa adoption studies. Am J Med Genet B Neuropsychiatr Genet 144(5):683-90). These scores and the rest of the available clinical data were then reviewed by two board-certified psychiatrists to provide two pools of individuals; a set of cases with severe, active ND and a set of age and ethnicity matched controls without a history of behavioral illness or significant alcohol, nicotine or illicit substance use.

[0166] The lymphoblast DNA used in the study was prepared from standard EBV transfected cell lines that were grown in standard bovine serum-based growth media supplemented with l-glutamine and penicillin-streptomycin as previously described (Philibert R A, Ryu G Y, Yoon J G, Sandhu H, Hollenbeck N, Gunter T, Barkhurst A, Adams W, Madan A. 2007. Transcriptional profiling of subjects from the Iowa adoption studies. Am J Med Genet B Neuropsychiatr Genet 144(5):683-90). The media was changed for each of these cell lines 24 hours prior to the extraction of DNA.

[0167] Input and methylation enriched fractions of DNA were prepared per the standard Nimblegen protocol (Roche Nimblegen I. 2007. Sample Preparation Protocol For DNA methylation Microarrays v3.0. Indianapolis). Briefly, 20 μg of DNA was reduced in complexity by digestion with Mse I, column, and a small aliquot taken for future analysis (i.e. input DNA). Five μg of the remainder of the digested DNA from each subject was resuspended in immunoprecipitation buffer (50 mM $NaPO_4$, 700 mM NaCl, 0.25% Triton X-100) and hybridized with 1 μg of monoclonal mouse anti-5-methyl cytidine antibody (Calbiochem USA) at 4° C. overnight. The resulting solution was then hybridized to a magnetic bead coupled secondary antibody (Dynabeads M-280, Invitrogen USA) and the DNA-antibody moiety purified by magnetic separation. The DNA was removed from the antibody complex by overnight digestion with protease K and column purified. Then 100 ng aliquots of both the methyl enriched DNA fraction and the input DNA were amplified using a WGA2 genome amplification kit used according to manufacturer's instructions (Sigma, St. Louis). After purification, this DNA was then frozen at −20° C. until use in the microarray analyses.

[0168] Hybridization to the 385K RefSeq whole genome promoter array (HG18 RefSeq) was conducted by Roche-Nimblegen (Indianapolis) under contract. These arrays contain 50-75 mer probes to 18,028 annotated RefSeq gene promoters with an average probe spacing of 100 bp. The resulting data, including the scaled $\log_2$ weighted ratios of the Cy3 (input) and Cy5 (methyl enriched) hybridization signals used in this report, were returned via courier.

[0169] The resulting data were then analyzed using a two-step process. In the first step, t-tests were conducted to identify probes whose hybridization values differed between the cases and controls at a significance level of $p<0.01$ (uncorrected). A clustering algorithm was then applied to this reduced probe set to identify probes which co-localized.

[0170] Bisulfite confirmation of differential methylation was conducted using standard procedures. Briefly, the DNA for each subject was first bisulfite modified then amplified using an Epitech® 96 Bisulfite and an Epitech® Whole Bisulfitome kit (both Qiagen, USA) according to manufacturer's instructions. The DNA samples were then amplified using a nested PCR protocol (1^{st} round primers, AGT GTT GGT GTA TTT ATT TTA AAA (SEQ ID NO:10) and TCC TAA AAA CAA ATA TCT TTC AAT C (SEQ ID NO:11); 2^{nd} round primers TAA CAA TAC TAA TCA TTT CAT AAA ATA (SEQ ID NO:12) and AGT TTA GTA ATT TGG AAT AAT AGG TTT (SEQ ID NO:13)). The resulting PCR products were gel purified, cloned using a StrataClone TA cloning kit (Stratagene USA), then sequenced at the University of Iowa DNA facility. The methylation status of each residue was then determined using CpG Viewer (Carr I M, Valleley E M A, Cordery S F, Markham A F, Bonthron D T. 2007. Sequence analysis and editing for bisulphite genomic sequencing projects. Nucl Acids Res 35(10):e79-) and the resulting data was analyzed via chi-square testing.

[0171] Gene pathway analysis was conducted using the web version of GOMiner™ using the default settings (Zeeberg B, Feng W, Wang G, Wang M, Fojo A, Sunshine M, Narasimhan S, Kane D, Reinhold W, Lababidi S and others. 2003. GoMiner: a resource for biological interpretation of genomic and proteomic data. Genome Biology 4(4):R28) while frequency analyses were conducted using the binomial test (Fleiss J L. 1981. Statistical Methods for Rates and Proportions. New York, N.Y.: John Wiley & Sons Inc.). Comparison of Cy3/Cy5 weighted values was conducted using logistic regression (Fleiss J L. 1981. Statistical Methods for Rates and Proportions. New York, N.Y.: John Wiley & Sons Inc.).

Results

[0172] The clinical and demographic information for the 41 subjects used in the case and control analyses are given in Table 8. All subjects were White with the average age of the cases being 43±7 years old and the controls 46±7 years old ($p<0.17$). The cases averaged over a pack of cigarettes per day at the time of phlebotomy with almost all of them having smoked heavily for over 20 years.

TABLE 8

Clinical and Demographic Data

	Cases		Controls	
	Male	Female	Male	Female
N	10	13	10	8
Age (years ± SD)	47 ± 8	41 ± 5	47 ± 9	46 ± 5
DSM IV ND Symptom Count	4.8 ± 2.0	5.2 ± 1.0	—	—
FTND	4.7 ± 2.7	4.4 ± 2.4	—	—
Daily Cigarette Consumption	25 ± 9	22 ± 9	—	—
Years Smoking	24 ± 10	21 ± 7	—	—

*Fagerstrom Test for Nicotine Dependence Scale (FTND), DSM IV Diagnostics and Statistics Manual Version 4

[0173] The methylation signals were analyzed as a group and by gender. The first analysis contrasted the signal from all cases versus all controls. The second analysis, consistent with prior strategies for analyzing behavioral data featured gender specific analyses.

[0174] As the initial step of all cases (n=23) vs controls (n=18) contrast, t-tests were conducted comparing the scaled and weighted Cy5/Cy3 ratios of the cases to that of the controls. Overall, the hybridization signal for 2534 probes differed at an uncorrected p-value of $p<0.01$. Because the clinical phenomenology associated with ND differs between males and females in our population, we then conducted gender specific analyses using the same methods. In this contrast, the male ND cases had 1790 probes that were differentially methylated at a p value of <0.01 while the female ND cases had 2070 probes that were differentially methylated at an uncorrected p value of $p<0.01$. Fifteen of the significant probes in the female only contrast were also found to be significant in the male only contrast ($p<0.03$) with two of those probes localizing to the same gene promoter (SLCO2B1).

[0175] Since our prior work in this area has demonstrated that methylation analyses are inherently noisy, we performed a cluster analysis of all significant probes from the combined set (<0.01) in order to increase the likelihood that the gene promoters selected for further analysis would represent real signal. The distribution of these differentially hybridizing probes was significantly nonrandom with 237 of the probes, localizing to just 113 gene promoters ($p<0.0001$). Seven gene promoters had three significant probes. Table 9 gives the HUGO approved names for the 106 genes that have names and which have two or more significant probes localizing to the gene promoter.

TABLE 9

List of Genes with Two or More Significant Probes

3 Significant Probes						
ANKRD13A	ATG2A	AX2R	CSNK1G2	NOVA1	SETBP1	SLMO2
2 Significant Probes						
AFAP1L1	ATP11A	C15orf57	C16orf61	C6orf195	C7orf45	C9orf72
CAMTA1	CARHSP1	CCDC144NL	CCNH	CCT6B	CFTR	CMIP
CNTD1	COL4A3	CSMD1	CTBP2	D2HGDH	DDX41	DLGAP2
DMRTA2	DOPEY2	EBF2	EIF4H	ELL	EMP3	ENTPD2
ENTPD2	EPOR	FER	FGR	FHDC1	FHOD1	FOXC1
FXN	GDF10	GFPT2	GPM6A	GRIK2	HIRIP3	HIST1H2BK
HIST2H2AA3	HSD17B4	IFNA17	ISL2	JPH2	KBTBD2	LAX1
LBX1	LOC254559	LRRC66	LRRN2	MAT2B	NECAB3	NID2

TABLE 9-continued

List of Genes with Two or More Significant Probes

NUBP1	OTOP1	PARP4	PDE5A	PNLDC1	PNMA5	PPIA
PRKAR1A	PRR7	PTDSS2	PTPRN2	RAC1	RBM20	RNPS1
RPIA	RPL39L	RPS17	SAV1	SCG5	SFRS17A	SGOL2
SH2D4B	SKP1	SLC25A21	SLC5A5	SLCO2B1	SOX17	SSTR1
STK40	TACR3	TBC1D8B	TESC	THOP1	TMC2	TOPORS
TP53INP1	ZIC5	ZNF148	ZNF830	ZPLD1		

The list of RefSeq genes with CpG islands containing two or more significantly probes that were symmetrically associated with active nicotine dependence ($p < 0.01$ nominal) in the genome wide analysis. Briefly, to generate this list, the normalized Log 2 hybridization ratios scores were analyzed a two-step process. In the first step, genome wide t-tests were conducted to identify probes whose hybridization values differed between the cases and controls at a significance level of $p < 0.01$ (uncorrected). A clustering algorithm was then applied to this reduced probe set to identify probes which co-localized to a 1000 bp sliding winding in the same island. Then, the genomic location of the CpG was checked against the HG 18 build of the human genome to identify RefSeq annotated genes associated with the island.

[0176] These 106 named genes from Table 9 were then subjected to pathway analysis using GOMiner™ (Zeeberg and others 2003) to identify gene pathways whose methylation patterns are differentially affected by smoking. In brief, the results of the analysis show that epigenetic change in proteins associated with cell proliferation and transmembrane transport are recurrent themes in these analyses (Table 10).

TABLE 10

Gene Pathway Analysis of the 113 Promoters with 2 or More Significant Probes.

Go Miner Category	Changed Genes/ Total Genes	P Value
GO:0007215 Glutamate Signaling Pathway	2/5	<0.001
GO:0008285 Negative Regulation of Cell Proliferation	6/128	<0.002
GO:0016607 Nuclear Speck	4/59	<0.002
GO:0016614 Oxidoreductase Activity Acting on CH—OH	4/60	<0.003
GO:0003007 Heart Morphogenesis	2/9	<0.003
GO:0000786 Nucleosome	2/11	<0.005
GO:0042626 ATPase Activity Coupled to Transmembrane Movement of Substances	2/11	<0.005
GO:0022804 Active Transmembrane Transporter Activity	3/37	<0.005
GO:0016820 Hydrolase Activity Acting on Acid Anhydrides	2/12	<0.006
GO:0043492 ATPase Activity Coupled to Movement of Substances	2/12	<0.006
GO:0022414 Reproductive Process	6/178	<0.006

TABLE 10-continued

Gene Pathway Analysis of the 113 Promoters with 2 or More Significant Probes.

Go Miner Category	Changed Genes/ Total Genes	P Value
GO:0016604 Nuclear Body	4/79	<0.006
GO:0007548 Sex Differentiation	3/45	<0.008
GO:0051082 Unfolded Protein Binding	3/45	<0.008
GO:0015276 Ligand-Gated Ion Channel Activity	2/15	<0.008
GO:0022834 Ligand-Gated Channel Activity	2/15	<0.008
GO:0000003 Reproduction	8/324	<0.009
GO:0030551 Cyclic Nucleotide Binding	2/16	<0.009
GO:0003006 Reproductive Developmental Process	3/49	<0.01
GO:0022892 Substrate-Specific Transporter Activity	7/265	<0.01

[0177] In order to validate the microarray analyses, we conducted sequencing of plasmid clones of bisulfite PCR products from 15 randomly selected cases (92 clones in total) and 15 randomly selected controls (77 clones in total) with respect to the AX2R promoter across the three significant probe regions identified in our initial analyses which localized to this gene. FIG. **8** shows the structure of the tiled region of the AX2R gene promoter. Table 11 gives the Cy3/Cy5 methylation ratios for the cases and controls, as well as their uncorrected p-values. When evaluating the strength of these p-values, it is important to note that these probes all recognize the same DNA contig produced by the Mse I digest.

TABLE 11

Sequence and Significance of AX2R Probes

	Cy3/Cy5 ratio (input to methyl enriched fraction)			
Probe Sequence	Avg Cases	Avg Controls	Difference	Pvalue*
ttcaggtgccaggtctggagtgctggtgcacctatctcaaaacgctgtct (SEQ ID NO: 14)	1.58	1.32	0.27	<0.03
gcaaacagcagtccagtaacctggaacaacaggctctgcgaaaccaagga (SEQ ID NO: 15)	1.84	1.48	0.36	<0.005
agaaatgaatggcgttgtcatcgaaaaaacacagactcgattgtgacagaaataccg (SEQ ID NO: 16)	1.66	1.34	0.32	<0.005
tgcgcctccacggaataactgccagccggcacagtgcgagtgagaaaccg (SEQ ID NO: 17)	1.74	1.40	0.34	<0.009

TABLE 11-continued

Sequence and Significance of AX2R Probes

Probe Sequence	Cy3/Cy5 ratio (input to methyl enriched fraction) Avg Cases	Avg Controls	Difference	Pvalue*
ggaaaagaatccgacgtcgccaacaagcggtgctaccaggagaaacgcct (SEQ ID NO: 18)	1.52	1.29	0.23	<0.09
aaaacacagctggataaaccgagaaccttcggagtggttgcaccgaaacg (SEQ ID NO: 19)	1.52	1.29	0.23	<0.09
gaagcaaccggcagtgctaacaccgaggagcacctagagcggcaaaacta (SEQ ID NO: 20)	1.18	0.86	0.32	<0.04

[0178] Table 12 gives the average methylation ratios for the sequenced CpG residues in the targeted region. As evidenced by the consistently elevated Cy3/Cy5 ratios across the promoter region, there was a relative decrease in the amount of methylation in the ND subjects as compared to the controls. This was particularly evident in the second, third and fourth probes covering the region. Bisulfite sequencing of this region of plasmid clones containing inserts from the PCR products of the bisulfite converted DNA samples from cases and controls confirmed those observations and demonstrated a nearly twofold greater amount of unmethylated residues in the smoking subjects as compared to the controls (average methylation in cases vs controls; 77.6% vs 88.8%, p<0.0001).

TABLE 12

Average Methylation at Bisulfite Sequenced Residues at AX2R

CG 4	CG 5	CG 6	CG 7	CG 8	CG 9	CG 10	CG 11	CG 12	CG 13	CG 14	CG 15	CG 16	CG 17
Cases*													
85%	77%	78%	71%	79%	78%	78%	76%	79%	80%	78%	71%	78%	77%
Controls													
88%	91%	87%	92%	88%	91%	87%	90%	92%	92%	88%	74%	94%	88%

Average number of residues successfully counted per CpG residue in the cases and controls was 74 and 89, respectively.

[0179] Finally, in order to compare our results with respect to previous published results using peripheral lymphocyte DNA (Launay J-M, Del Pino M, Chironi G, Callebert J, Peoc'h K, Megnien J-L, Mallet J, Simon A, Rendu F. 2009. Smoking Induces Long-Lasting Effects through a Monoamine-Oxidase Epigenetic Regulation. PLoS ONE 4(11): e7959), we compared the probe values between cases and controls at this X-chromosome locus. Consistent with prior findings, the amount of methylation at the MAOB promoter was significantly decreased in both males (LR; p<0.006) and females (LR; p<0.007).

Discussion

[0180] In summary, it is reported that smoking is associated with both altered overall and locus specific alterations in DNA methylation with particular enrichment of altered methylation in pathways associated with glutamate signaling, cell proliferation and detoxification. Strengths of this study are the well characterized subjects, the similarity of the whole genome promoter array results for males and females, and the sequencing confirmation.

[0181] The vast majority of the loci with differential methylation in this study are not directly involved with neurotransmission. Consistent with the role of smoking in cancer and altered DNA methylation part of the oncogenic process, it is logical to find that each of the 7 genes with 3 significant probes have suggested roles in carcinogenesis (Buckanovich R J, Yang Y Y, Darnell R B. 1996. The onconeural antigen Nova-1 is a neuron-specific RNA-binding protein, the activity of which is inhibited by paraneoplastic antibodies. J Neurosci 16(3):1114-1122; Greenman C, Stephens P, Smith R, Dalgliesh G L, Hunter C, Bignell G, Davies H, Teague J, Butler A, Stevens C and others. 2007. Patterns of somatic mutation in human cancer genomes. Nature 446(7132):153-8; Koike Folgueira M A, Brentani H, Carraro D M, De Camargo Barros Filho M, Hirata Katayama M L, Santana de Abreu A P, Mantovani Barbosa E, De Oliveira C T, Patrao D F, Mota L D and others. 2009. Gene expression profile of residual breast cancer after doxorubicin and cyclophosphamide neoadjuvant chemotherapy. Oncol Rep 22(4):805-13; Ma S, Huang J K, Shen S. 2009. Identification of Cancer Associated Gene Clusters and Genes Via Clustering Penalization. Statistics and Its Interface 2:1-11; Masayoshi M, Naoki K, Manuel J G-R, Tatsuo A, Terry D C, Kunihiro U, Yoshifumi A. 2001. Identification and characterization of SEB, a novel protein that binds to the acute undifferentiated leukemia-associated protein SET. European Journal of Biochemistry 268(5):1340-1351; Wright P K, May F E, Darby S, Saif R, Lennard T W, Westley B R. 2009. Estrogen Regulates Vesicle Trafficking Gene Expression in EFF-3, EFM-19 and MCF-7 Breast Can-

cer Cells. Int J Clin Exp Pathol 2(5):463-75; Yusuke S, Aaron M H, Younghun J, Anne M Z, Elisabeth A P, Jingcheng W, Jianhua W, Ganwei L, Roodman G D, Robert D L and others. 2008 Annexin II/Annexin II receptor axis regulates adhesion, migration, homing, and growth of prostate cancer. p 370-380). This suggests that the processes affected by smoking in other cells may be reflected in the differential methylation of the lymphoblasts, and that the lymphoblast model may provide a reasonable representation of systemic methylation changes.

[0182] In light of this apparent enrichment of genes involved carcinogenesis, it is notable that most significant Gene Ontology (Gene Ontology C. 2004. The Gene Ontology (GO) database and informatics resource. Nucl Acids Res 32 (suppl_1):D258-261) pathway identified in the GOMiner™ analysis in this study is the glutamate signaling pathway (on the basis of GRIK2 and SSTR1). GRIK2 gene expression is decreased in the brains of smoking mice (Wang J, Gutala R, Hwang Y, Kim J, Konu O, Ma J, Li M. 2008. Strain- and region-specific gene expression profiles in mouse brain in response to chronic nicotine treatment. p 78-87) and genetic variation in GRIK2 (Vink J M, Smit A B, de Geus E J, Sullivan P, Willemsen G, Hottenga J J, Smit J H, Hoogendijk W J, Zitman F G, Peltonen L and others. 2009. Genome-wide association study of smoking initiation and current smoking Am J Hum Genet 84(3):367-79) was linked to smoking in recently published GWAS of smoking. These recent and other prior findings support a role for glutamate signaling in the mood altering and drug reinforcing effects of nicotine (Lambe E K, George T P. 2008. Perspective: Translational Studies on Glutamate and Dopamine Neurocircuitry in Addictions: Implications for Addiction Treatment. Neuropsychopharmacology 34(2):255-256). These current results add to that body of evidence and further suggest that the role of glutamate signaling system should receive greater attention in analyses of mechanisms of addiction associated with smoking.

[0183] It will be important to identify which of these methylations are static and which are dynamic. In our previous work at MAOA, we found that reduction in methylation was particularly pronounced as a result of smoking cessation and given MAOA's prominence in catabolizing dopamine, we speculated that this epigenetic change could be part of the withdrawal syndrome. Given the current systematic findings, the findings of others with respect to methylation of the MAOB gene promoter (Launay J-M, Del Pino M, Chironi G, Callebert J, Peoc'h K, Megnien J-L, Mallet J, Simon A, Rendu F. 2009. Smoking Induces Long-Lasting Effects through a Monoamine-Oxidase Epigenetic Regulation. PLoS ONE 4(11):e7959) and our results at MAOB, it is unlikely that the MAOA promoter is the only regulatory motif changed after smoking cessation. If so, by studying withdrawal on a genome wide basis, it may be possible to more readily identify the pathways involved in nicotine craving and devise more effective interventions to short circuit this disruptive syndrome that obfuscates effective treatment.

Example 5

Methylation Profiling of Alcohol Dependence

[0184] The list of RefSeq genes with CpG islands containing two or more significantly probes that were symmetrically associated with active alcohol dependence ($p<0.001$ nominal) in the genome wide analysis is provided in Table 13. Briefly, to generate this list, the normalized Log 2 hybridization ratios scores were analyzed a two-step process. In the first step, genome wide t-tests were conducted to identify probes whose hybridization values differed between the alcohol cases and controls at a significance level of $p<0.001$ (uncorrected). A clustering algorithm was then applied to this reduced probe set to identify probes which co-localized to a 1000 bp sliding winding in the same island. Then, the genomic location of the CpG was checked against the HG 18 build of the human genome to identify RefSeq annotated genes associated with the island.

TABLE 13

ABCA12	EMILIN3	ODZ2	WAC
ABL2	EXOC6B	ODZ4	WASF2
AGBL1	FAM125B	PARK2	WDR78
AK097539	FBXL4	PDAP1	WSCD2
AK125749	FCGBP	PDGFA	XPR1
AK128353	FLJ16779	PDGFRA	XRCC5
AK129763	FOXN3	PDLIM1	ZBTB7B
AK309744	FXR1	PDXDC2	ZDHHC2
AK311380	GAD2	PDZD2	ZFHX4
AKAP12	GNB3	PEX7	ZFP92
AMPH	GRAP2	PIH1D1	ZNF221
ANKRD53	GTF2I	PLSCR3	ZNF263
APBA1	HCCA2	PNRC1	ZNF33A
APBB2	HEXIM1	PON2	ZNF423
ARHGAP10	HEXIM2	PPM1A	ZNF623
ARHGEF16	HIP1R	PSG6	ZSCAN5A
ARL17	HISPPD1	QSER1	
ATP6V1E1	HOXA2	RAB26	
BC032407	HSF1	RMND5A	
BC051727	IMAA	ROS1	
C11orf64	IMMT	RPIA	
C12orf53	INTS10	RSPO1	
C1orf101	ITGA5	RUNDC2C	
C20orf117	KCNH5	SCT	
C7orf50	KCNQ1	SDF4	
C9orf72	KLHDC1	SLC02B1	
C9orf82	KLHL9	SMCR7L	
CCBL1	LHFPL3	SNTB2	
CDC123	LNPEP	SPATA5	
CDH5	LOC100133545	SPDYE3	
CHR11:002610294	LOC284805	SPNS2	
COL2A1	LRSAM1	SRL	
COPS7A	MECP2	STAM2	
CPNE4	MIB2	STK36	
CR936796	MLL	SYT13	
CSMD1	MPZL1	TANC1	
CSRNP3	MYO9B	TJP2	
DGKH	NAT9	TMEM205	
DNHD1	NBPF14	TNRC6B	
DOCK11	NEAT1	TRAF3IP2	
DOCK4	NF2	TXNDC11	
DPY19L4	NME2P1	USP45	
DYNC1LI1	NOC2L	UVRAG	
EDEM2	NPAS1	VCPIP1	

Example 6

Methylation Profiling of Cannabis Dependence

[0185] Table 14 provides a list of CpG residues whose methylation was significantly associated with Cannabis Dependences at a nominal p-value of $p<0.05$. For female subjects: CpG 69 and CpG 88. For male subjects: CpG 11-12, 13, 64, 69, 72-73. Unpublished data from Philibert et al., 2008 "MAOA methylation is associated with nicotine and alcohol dependence in women."

[0186] The list of RefSeq genes with CpG islands containing two or more significantly probes that were symmetrically associated with active cannabis dependence ($p<0.01$ nomi-

nal) in the genome wide analysis is provided in Table 14. Briefly, to generate this list, the normalized Log 2 hybridization ratios scores were analyzed a two-step process. In the first step, genome wide t-tests were conducted to identify probes whose hybridization values differed between the cannabis cases and controls at a significance level of $p<0.01$ (uncorrected). A clustering algorithm was then applied to this reduced probe set to identify probes which co-localized to a 1000 bp sliding winding in the same island. Then, the genomic location of the CpG was checked against the HG 18 build of the human genome to identify RefSeq annotated genes associated with the island.

TABLE 14

AK056486	ANKHD1	LOC283050	FNTB	SRRD
HES4	RNF14	GSTO2	WDR25	PATZ1
BC033949	EBF1	BC132944	AKT1	DNAL4
SDF4	RREB1	PWWP2B	TMCO5A	CYP2D7P1
AURKAIP1	TXNDC5	DRD4	MGA	MIOX
AX747988	ABT1	PNPLA2	CDAN1	SHOX
MSTP2	GLP1R	HCCA2	TSPAN3	CD99
ECE1	C6orf130	HCCA2	C16orf13	RPL39
C1orf212	TRERF1	AK126380	TMEM159	MCF2
TEKT2	TAAR1	IGF2	ATP2A1	
STK40	UST	INS-IGF2	IMAA	
CYP4Z1	AKAP12	TRPM5	IRX3	
PARS2	FNDC1	KCNQ1	RTN4RL1	
NBPF16	IGF2R	KCNQ1	TEKT1	
MSTO1	IGF2R	KCNQ1	SLC25A35	
KIAA0907	BC087858	SLC22A18AS	GAS7	
TOMM40L	AK299216	NAP1L4	HS3ST3B1	
PTGS2	HOXA	OSBPL5	C17orf76	
AK095633	AK093987	MRGPRE	USP22	
OR2T1	CDK13	DENND5A	STAT5B	
MYT1L	C7orf40	CALCB	KPNB1	
TSSC1	GTF2IRD1	SYT13	TMEM100	
AK055918	LMTK2	APLNR	AXIN2	
EPAS1	SLC26A5	MACROD1	ASPSCR1	
TMEM177	C7orf60	KCNK4	TBCD	
GAL3ST2	FLJ43663	EHD1	SETBP1	
EFHB	EXOC4	PPP2R5B	ST8SIA5	
GLT8D1	MFHAS1	CAPN1	LIPG	
GLT8D1	NUDT18	PITPNM1	CTDP1	
ITIH4	UNC5D	PITPNM1	BSG	
KIAA1013	DPY19L4	GPR83	GPX4	
PDZRN3	LY6K	NCAM1	SBNO2	
CGGBP1	DNAJB5	OPCML	STK11	
RG9MTD1	UNC13B	IFFO1	KIAA1532	
IGSF11	FXN	SLC2A3	ALKBH7	
AMOTL2	C9orf85	NDUFA4L2	ICAM1	
GNB4	PCSK5	PEBP1	KCNA7	
RPL39L	AK309476	SIRT4	ETFB	
MFSD7	MEGF9	OASL	ZNF530	
AX748388	CIZ1	RSRC2	TCF15	
CRMP1	USP20	RIMBP2	PSMF1	
PDGFRA	NTNG2	IFT88	BTBD3	
H2AFZ	GTF3C5	PARP4	SLC12A5	
LARP7	CAMSAP1	ESD	EYA2	
CLCN3	LCNL1	TPPP2	ZNF217	
ANKRD37	C10orf18	EFS	TPD52L2	
AHRR	STOX1	C14orf147	DNAJC5	
IRX1	CHST3	SOCS4	SIK1	

Example 7

Dose Dependent Impact of Recent Alcohol Use on Genome-Wide DNA Methylation Signatures

[0187] Together, alcohol use and dependence affect 8% of the adult United States each year and cause over 200 billion dollars of economic damage annually. The mechanism(s) through which alcohol exerts this toll varies. During acute intoxication, much of the economic damage and personal injury results from the increased rate of accidental injury. But after returning to sobriety, the risk for further damage from accidental injury markedly diminishes. However, in the case of the sustained heavy use of alcohol, the risk for increased morbidity does not remit after return to sobriety and the individual remains at increased risk for a large number of medical conditions including hypertension, heart disease, and impaired executive function in the absence of acute intoxication. At the microscopic level, this increased risk can be directly linked to adverse impact on tissue and organ damage. However, at the molecular level, the direct effects of long term alcohol use seem more complex with chronic changes in a number of biochemical pathways being noted.

[0188] Some of these cellular changes may be legacies of altered protein folding and trafficking bequeathed to the cell from periods of intoxication. However, most cellular proteins have limited lifetimes before they are intracellularly recycled. Hence, they are unlikely to be directly responsible for some of the chronic dysfunction seen in cells prepared from abstinent alcoholics. Instead, some of these alterations may result from alcohol induced changes in genomic tone, which is defined as the stable transcriptional repertoire of a cell.

[0189] The factors that control the "genomic tone" or transcriptional repertoire of the given cell are diverse but can be generally categorized as genetic variation, tissue specific transcriptional activators/repressors and epigenetic factors. Conceivably, chronic alcohol use could affect the type and distribution of both transcriptional and epigenetic factors thus changing the genomic tone of the given cell. Unfortunately, systematic methods for assessing tissue specific transcription factors are not commonly available. In contrast, recent advances in DNA methylation assessment technologies have made genome wide assessment of DNA methylation more accessible.

[0190] This advancement is particularly welcome because in prior work using more restricted approaches, we and others have presented evidence that alterations in DNA methylation may be in part responsible for altered genomic tone observed in peripheral blood cells from subjects who chronically use alcohol. However, these studies were limited by the low number of genes surveyed and the limited number of subjects surveyed. In this communication, using the Illumina HumanMethylation450 BeadChip, which interrogates over 485,000 CpG residues, we examine the relationship between alcohol consumption and degree of DNA methylation in lymphoblast DNA prepared from 165 female subjects from the Iowa Adoption Studies, the largest case and control adoption studies of substance use in the world.

Methods

[0191] The protocols and procedures used in the Iowa Adoptions Studies (IAS) have been described in detail elsewhere (Yates et al., 1998, The Iowa Adoption Studies Methods and Results; In: LaBuda et al, Ed's., *On the Way to Individuality: Methodological Issues in Behavioral Genetics*, Hauppauge N.Y.; Nova Science Publishers, pp 95-125). In brief, the IAS is a case and control adoption study of the effects of genetic, environmental and gene-environment interactions in the etiology of substance use and antisocial personality. The data used in the current study is derived from interviews with the Semi-Structured Interview for the Assessment of the Genetics of Alcoholism, Version II (Bucholz et al., 1994, A new, semi-structured psychiatric interview for use in genetic linkage studies; a report on the reliability of the

SSAGA, *J. Stud. Alcohol,* 55:149-58), during each of the last two waves of the IAS study (1999-2004 and 2005-2009). Using this data, subjects were classified on the basis of their alcohol use in the past six months prior to assessment into four categories: 1) abstinent (no use in the past six months); 2) mild users (use of alcohol in between 1-8 weeks in the past 6 months); 3) moderate users (use of alcohol in between 9 and 25 weeks in the past six months); and 4) heavy users (alcohol use in every week in the past six months). The lymphoblast DNA was derived by Ebstein Barr virus mediated transformation (Caputo et al., 1991, An Effective Method for Establishing Human B Lymphoblastic Cell Lines Using Epstein Barr Virus, *J. Tiss. Cult. Meth.,* 13:39-44) of lymphocytes obtained from blood donated by 165 female subjects during the last wave of the study.

[0192] The lymphoblast DNA used in this study was prepared from growth-entrained lymphoblast cell lines using our standard procedures (Philibert et al., 2008, MAOA methylation is associated with nicotine and alcohol dependence in women, *Am. J. Med. Genet. B. Neuropsychiatr. Genet.,* 147: 565-70). In brief, on the day before DNA preparation, one-half of the cell media for each culture flask was exchanged. DNA was then prepared from the cell lines twenty four hours later using cold protein precipitation (Lahiri et al., 1993, DNA isolation by a rapid method from human blood samples: effects of $MgCl_2$, EDTA, storage time, and temperature on DNA yield and quality, *Biochem. Genet.,* 31:321-8). After quantification and purity assessment using a Nanodrop (Thermo Scientific, USA) spectrophotometer, DNA was stored at −20° C. and RNA was −80 C until use.

[0193] Genome wide DNA methylation of the DNA was assessed using the Illumina HumanMethylation450 BeadChip under contract by the University of Minnesota Genome Center using the protocol specified by the manufacturer and the contractor. The resulting microarray data were inspected for complete bisulfite conversion of the DNA, and average beta values (i.e. average methylation) for each CpG residue were determined using the GenomeStudio V2009.2; Methylation module Version 1.5.5, version 3.2 (Illumina, San Diego). The resulting beta values were exported into Microsoft Excel and JMP (SAS Institute, USA) for data analysis. The HumanMethylation450 BeadChip contains 485,577 probes that recognize at least 20216 unique features. With respect to this sample, >99.7% of the 485,577 probes yielded statistically reliable data.

[0194] The methylation status of the serotonin transporter (SLC6A4) promoter region was previously assessed for 163 of the 165 samples using MALDI-TOF mass spectroscopy by Sequenom (San Diego, Calif.) as described previously (Philibert et al., 2008, MAOA methylation is associated with nicotine and alcohol dependence in women, *Am. J. Med. Genet. B. Neuropsychiatr. Genet.,* 147:565-70). Using the sequence annotation files from both the current and the prior studies, we identified CpG residues that were assessed using both technologies. The methylation values for each residue were compared using Least Squares regression.

[0195] After logarithmic conversion, data were inspected for outliers and the initial data analyses were conducted using genome wide t-tests. Subsequently, beta values for each of the probes were aligned according to their physical location and the data re-analyzed using paired t-tests over a 11-probe sliding window in order to more adroitly capture methylation signatures over larger regions (Dindot et al., 2009, Epigenetic profiling at mouse imprinted gene clusters reveals novel epigenetic and genetic features at differentially methylated regions, *Genome Res.,* 19:1374-83; Farthing et al., 2008, Global Mapping of DNA Methylation in Mouse Promoters Reveals Epigenetic Reprogramming of Pluripotency Genes, *PLoS Genet.,* 4:e100-16). All genome wide comparisons were corrected for multiple comparisons using the method of Benjamini and Hochberg (1995, Controlling the false discovery rate: a practical and powerful approach to multiple testing, *J. Royal Statist. Soc., Series B. Methodological,* 57:289-300). For select loci, data were analyzed with respect to alcohol use status using ANOVA.

[0196] Pathway analysis of differentially methylated genes was conducted using GoMiner™ using default settings (Zeeberg et al., 2003, GoMiner: a resource for biological interpretation of genomic and proteomic data, *Genome Biol.,* 4:1-8). All values reported include nominal and FDR corrected values.

Results

[0197] The demographic and clinical characteristics of the 165 female subjects are shown in Table 15. Overall, the subjects were largely white and tended to be their mid-to-late 40s. Consistent with enrichment of the sample for the diathesis of substance use, the majority of the subjects in the study reported the use of alcohol in the past six months with the 28 subjects in the "heavy" use group reporting alcohol use every week for the past 26 weeks while the 50 "moderate" drinkers and the 47 "mild" drinkers reported drinking in 9-25 weeks, and 1 to 8 weeks in the past 26 weeks, respectively. Their current drinking pattern was reflective of their lifetime history of drinking alcohol. Only 5 of the 40 individuals who reported recent abstinence also reported 1 or more symptoms of lifetime alcohol dependence. In contrast, 19 of 28 of the heavy drinking reported one or more symptoms ($p<0.01$) with 7 of them meeting criteria for a lifetime diagnosis of alcohol dependence (3 or more symptoms). Approximately 50% of the subjects also reported a past or former history of smoking with 27% continuing to smoke at the time of phlebotomy. However, despite the strong epidemiological associations of smoking and drinking behaviors, in this cohort of 165 subjects, there were no significant differences in the rates of smoking between the three groups.

[0198] In our initial analyses, we contrasted the methylation values for the 40 abstinent individuals with the values for the 47 mild, 50 moderate and 28 heavy drinkers using genome wide t-tests. The results of those analyses are shown in Table 16. As the table indicates, although some of the values show strong consistency across several partially independent comparisons, by themselves none of the comparisons between individual groups (e.g. heavy drinker vs. mild) are statistically significant after genome wide comparison (best p-value after correction is $p<0.25$). However, when the moderate and heavy drinkers are pooled together, the comparison at cg05600126, a probe in ABR, a gene known to be involved in vestibular function (Kaartinen et al., 2002, Vestibular dysgenesis in mice lacking Abr and Bcr Cdc42/RacGAPs, *Develop. Dynamics,* 223:517-25), reaches genome wide significance after genome wide correction ($p<0.05$; Table 17) with several other probes nearly reaching significance.

[0199] There appears to be a dose dependent effect of weeks of drinking on both the strength of the overall comparisons (Table 18). Overall, 1711 of the 485,577 probes on the array were nominally significant at the $p<0.001$ level in the heavy vs. abstinent group comparison (expected value;

486 probes at p<0.001). This number diminishes to 390 probes when comparing the moderate to the abstinent and 128 probes at the p<0.001 level when comparing the mild drinkers to the abstinent drinkers despite the fact that the heavy drinker group was the smallest of the three groups.

[0200] This effect is also reflected in the distribution of the differentially methylated probes with respect to their island status. In previous work, it was shown that changes in cell fate preferentially affected CpG methylation based on the location of the residue with respect to their location in the CpG island. To examine whether this was happening with respect to alcohol use, we examined the location of the significantly differentially methylated probes (at the p<0.001 level) using the information contained in the Illumina file annotations. As Table 18 demonstrates, alcohol seems to preferentially affect the probes found in the center of the CpG islands with the proportion of all differentially methylated probes which localized to the center of the islands rising reaching 52% (p<0.0001) in the heavy drinking group.

[0201] Next, using a sliding 11 probe window, we examined whether using information from adjacent probes would strengthen the findings with respect with to alcohol use. The effect on significance was profound with values for 19 regional comparisons reaching genome wide significance at the p<0.001 level (Table 19). Not surprisingly, many of the regions are overlapping with the top 4 region comparisons all being found in BLCAP, a chromosome 6 gene with 121 probes localizing to it.

[0202] Using the 1711 probes nominally differentiated at the p<0.001 level, we conducted pathway analyses using GoMiner. As Table 20 indicates, pathways that involve large networks of genes, in particular those affecting basic nucleic acid and cellular metabolic processes, were strongly affected which suggests that the effect of alcohol consumption over the most recent 6 months on gene methylation are widespread and not limited to small, circumscribed pathways.

[0203] Finally, to determine whether our current array based measurements were valid, we compared the degree of methylation determined by the Illumina platform with the values determine previously for these 165 subjects at the serotonin transporter promoter associated CpG island (SLC6A4) using MALDI-TOF mass spectrophotometer. Overall, 4 CpG residues at this locus were surveyed by both approaches. At each CpG residue, the degree of methylation determined by each method was correlated with the average adjusted r2 equaling ~0.34, which strongly suggests that the current measurements are reliable.

Discussion

[0204] In summary, we demonstrate that the pattern of alcohol use over the most recent 6 months is associated with widespread changes in the methylation of lymphoblast DNA derived from middle aged women. These changes are modest, but widespread, affecting a broad portfolio of cellular metabolic processes. Limitations of the current findings include the fact that lymphoblasts are not primary human cells, the modest degree of differential methylation observed at any individual probe and the likely confounding effect of prior alcohol use history on six month use history. Strengths of the manuscript include the high significant multipoint analyses, the internal consistency of the multiple comparisons and the independent verification of methylation signatures at the SLC6A4 locus.

[0205] The "dose" dependency of differential DNA methylation observed in the current study was to be expected. Depending on context, alcohol can be viewed as either as a drug or a solvent which makes conceptualizing the observed epigenetic changes described herein as gradated responses to a cellular toxin natural. However, it is important to realize the primary alcohol use variable employed in the current study was number of weeks in the past six months in which the subject drank. To a certain extent, our choice to employ this measure is because our diagnostic instrument, the SSAGA, readily provides this as a recent use metric. It may well be that the choice of a different metric, such as the number of alcoholic beverages consumed in the past two weeks, may have produced more robust findings. However, because of the manner in which the alcohol use questions are asked in our version of this instrument, information for all subjects is not always available or directly comparable. Hence, it may well be that other approaches to quantifying recent alcohol consumption may produce more robust findings. But we feel that the current classification system which captures the pattern of use over an extended period of time may be equally effective and that further analyses using replicate data sets using more complex (e.g. factor analyses) may be the best way to more adroitly define which recent alcohol use measures are most correlated with DNA methylation changes.

[0206] The stepwise effect of alcohol use severity on the distribution of the differentially methylated probes with respect to CpG island status is intriguing. In previous work, it was demonstrated that DNA methylation changes associated with assumption of cell fate preferentially affected the less dense outer area of the CpG islands referred to, poetically, as the "shore" (Doi et al., 2009, Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts, *Nat. Genet.,* 41:1350-3). In this survey of the effects of alcohol intake, we observed that overall that greater consumption of alcohol is associated with increased levels of genome wide methylation and that the changes in the most chronically exposed subjects preferentially affects the centers of these islands. Because two thirds of all CpG islands in the genome are promoter associated and hypermethylation of promoter-associated CpG islands is thought to silence gene transcription, it is tempting to speculate that this increased methylation observed at these islands is associated with decreased gene expression at these loci. Unfortunately, because the magnitude of many of these changes in methylation are relatively small and others have observed that the relationship between DNA methylation and gene expression may be complex and weak, directly demonstrating that these changes have biological relevance at any given locus may be difficult. However, the finding that the expression of BLCAP, a gene region that is significantly hypermethylated in our study, was significantly decreased in the nucleus accumbens of ethanol treated rats is encouraging (Rodd et al., 2008, Differential gene expression in the nucleus accumbems with ethanol self-administration in inbred alcohol-preferring rats, *Pharmacol. Biochem. Behavior,* 89:481-98).

[0207] The directionality of the overall changes observed herein is consistent with prior findings. In our single point and sliding window analyses, almost of all of the top thirty most significantly differentially methylated probes or regions were more methylated in heavy alcohol use group. This is very consistent with prior finding by ourselves and others. However, there are a number of exceptions to this rule in this study

and we expect that a fuller understanding of differentially regulated pathways in alcoholism to include a rich tapestry of both up-regulated and down-regulated genes.

[0208] The results of the gene pathway analyses may seem surprising at first glance. Because many prior investigations of effects of alcohol have focused on the CNS, there is a strong bias in the extant literature with respect to CNS relevant pathways. However, in our pathway analyses, none of the most significant pathways implicated in the current study directly concern neurotransmission. Instead, the most differentially methylated pathways identified in the current study concern general cell metabolic functions such as membrane trafficking and nucleotide synthesis. This may be because that once ingested, alcohol evenly disperses itself through all tissue and although it may affect neurotransmission systems, such as the GABAergic system more prominently than others, none of these are directly receptor mediated effects. Instead, many of the most profound effects of acute alcohol ingestion on somatic function derive from the effect of alcohol on membrane polarity and the isoelectric properties of the cytosolic environment. Because these properties of the membrane and cytosolic compartments are so critical to cellular homeostasis, the processes found in these regions are those which are most affected in the pathway analyses.

[0209] In this regard, it is also important to note that despite the fact that alcohol and nicotine use are frequently co-morbid, there were no differences in the frequency of smoking between the four alcohol use groups and controlling for smoking had no effect on the outcomes of the current study. Taken in conjunction with previous findings, these findings suggest that cigarette smoke and alcohol ingestion present unique toxicological challenges to cells that have distinct effects on methylation.

[0210] A critical question that is not addressed by the current study is the longevity of the methylation signatures associated with chronic alcohol use. In prior studies of the MAOA locus, we have demonstrated that cessation of smoking has dramatic effects on CpG methylation. Unfortunately, the number of abstinent or nearly abstinent subjects contained within the current study is too small to conduct meaningful tests at the most significant loci for these purposes. Furthermore, not all abstinent individuals in this study were abstinent for the same reasons. Some are abstinent secondary to personal choice while others in our study are abstinent secondary to medical or legal necessity. Controlling for those and other potential confounders such as diet and lifestyle issues in small samples such as this may be difficult.

[0211] Assuming that the current findings are replicated, particularly in primary lymphocytes, some of the most critical questions to be addressed concern the relationship of differential DNA methylation to the overall genomic tone of the cell. DNA methylation is assumed to be intimately involved in regulation of genomic tone. Hence, will reversal of the DNA methylation changes restore normal genomic tone? This is an important question because cells isolated from alcoholics also have more structural changes such as shorter telomeres and manifest other signs of cellular senescence such as abnormal post-translational modifications of proteins. Will these indicators of cellular dysfunction similarly revert if the methylation patterns can be reversed through dietary or pharmacological means? If so, defining the methods through which to accomplish this process could have substantial impact in the rehabilitation of those suffering from the mental and physical ravages of alcoholism.

[0212] In summary, we report that recent chronic alcohol intake is associated with significant changes in CpG methylation, and in particular, increased hypermethylation of CpG islands. We suggest further studies to confirm and extend these findings using primary cells and convergent epigenetic approaches are indicated.

TABLE 15

Clinical Characteristics of the 165 Female Iowa Adoptions Studies Probands

		DRINKING STATUS			
		Abstinent	Mild	Moderate	Heavy
N		40	47	50	28
Age		47 ± 8	46 ± 8	44 ± 8	46 ± 8
Ethnicity					
White		38	45	49	26
Other		2	2	1	2
Smoking Status					
Current Smoker		7	9	13	9
Former Smoker		11	11	15	9
Never		22	27	22	10
Lifetime DSM IV Alcohol Dependence Symptom Counts					
Sxs	0	35	28	31	9
	1	4	9	8	7
	2	4	5	5	5
	3	3	2	2	3
	4	1	1	2	4
	5	0	0	1	0
	6	1	0	1	0
	7	1	1	0	0

TABLE 16

The Top 30 Most Significantly Associated Probes for Individual Alcohol Group Comparison.

				Average Methylation for each use group				Nominal p-values for group comparisons		
Probe ID	GENE	Placement	Island Status	Abs	Mild	Mod	Heavy	Heavy vs Abs	Mod vs Abs	Heavy vs Mild
cg24023553			N Shore	0.10	0.11	0.11	0.12	2.64E−06	0.0021	0.0280
cg20310749	SHC4	TSS1500	S Shore	0.05	0.06	0.06	0.07	2.68E−06	0.0068	0.0008
cg23865067	ARPP19	Body	N Shore	0.08	0.08	0.08	0.09	3.58E−06	0.0080	0.1053
cg05559557			Island	0.89	0.90	0.90	0.90	3.75E−06	0.0565	0.0042
cg24268236	CEP63	TSS200	Island	0.09	0.09	0.09	0.10	3.98E−06	0.0014	0.5055
cg09966309	RPS6KA2	Body		0.27	0.22	0.21	0.15	5.85E−06	0.0026	0.0510

TABLE 16-continued

The Top 30 Most Significantly Associated Probes for Individual Alcohol Group Comparison.

				Average Methylation for each use group				Nominal p-values for group comparisons		
Probe ID	GENE	Placement	Island Status	Abs	Mild	Mod	Heavy	Heavy vs Abs	Mod vs Abs	Heavy vs Mild
cg23818046	CENPK	TSS1500	Island	0.06	0.07	0.07	0.07	6.06E−06	0.0005	0.0774
cg22640209	DOCK10	Body	N Shore	0.05	0.06	0.06	0.06	6.92E−06	0.0023	0.0045
cg05128246	KHDRBS3	Body	Island	0.04	0.04	0.04	0.05	7.61E−06	0.0029	0.0682
cg07211915	MAP3K15	TSS200	Island	0.44	0.46	0.45	0.48	7.61E−06	0.0017	0.1484
cg02606081	HRAS	TSS1500	Island	0.13	0.13	0.14	0.15	8.47E−06	0.0026	0.0085
cg12502823	MGC70857	Body	Island	0.06	0.07	0.07	0.08	8.86E−06	0.0007	0.1258
cg05497240	C5orf4	3'UTR		0.83	0.84	0.85	0.85	8.89E−06	0.0038	0.0005
cg26248486	BBS10	TSS200S	Shore	0.07	0.07	0.07	0.07	1.00E−05	0.0249	0.0003
cg26213873	CTTNBP2NL	5'UTR	Island	0.10	0.11	0.12	0.12	1.03E−05	0.0427	0.0000
cg16480634	ACTR2	3'UTR		0.61	0.65	0.62	0.68	1.04E−05	0.1278	0.5470
cg17879912	TNFAIP8	Body		0.77	0.78	0.79	0.80	1.11E−05	0.0034	0.0049
cg23554129			Island	0.11	0.11	0.12	0.12	1.13E−05	0.0024	0.0019
cg00717297	TMEM120B	Body		0.85	0.86	0.86	0.87	1.14E−05	0.0001	0.0055
cg12999103	ATP13A2	Body	Island	0.12	0.12	0.12	0.13	1.28E−05	0.0004	0.0409
cg16551665	CDK5R1	TSS200	Island	0.12	0.12	0.12	0.13	1.32E−05	0.0006	0.0130
cg03461296	TAF4	Body	N Shore	0.83	0.84	0.84	0.85	1.36E−05	0.0246	0.0041
cg12361155	ADO	TSS200	Island	0.07	0.07	0.07	0.08	1.43E−05	0.0366	0.0367
cg25253419	NUCB1	TSS200S	Shore	0.09	0.10	0.10	0.11	1.46E−05	0.0172	0.0064
cg18634443	TBPL1	3'UTR		0.63	0.66	0.66	0.67	1.54E−05	0.3750	0.0003
cg05353415	GLI3	5'UTR	Island	0.07	0.08	0.08	0.08	1.58E−05	0.6627	0.0053
cg02988255	GPR44	Body	Island	0.78	0.79	0.80	0.81	1.61E−05	0.0014	0.0013
cg05944623	PRPF31	5'UTR	Island	0.08	0.08	0.08	0.09	1.62E−05	0.0001	0.0079
cg01766534	MRPL44	1stExon	Island	0.08	0.08	0.08	0.09	1.64E−05	0.0025	0.0320
cg15090909	TBC1D9B	Body	Island	0.86	0.87	0.87	0.88	1.69E−05	0.0022	0.0006

Abbreviations: Abs = abstinent, Mod = moderate, S = south, and N = north.

All methylation values are average beta values.

TABLE 17

The Top 30 Most Significantly Associated Probes in the Abstinent vs Pooled Moderate and Heavy Drinkers Analysis

				Average Methylation						BH Corrected
Probe ID	GENE	Placement	Island Status	Abst	Mod	Heavy	H vs Abs	Mod vs Abs	H&Mod vs Abs	Value
cg05600126	ABR	Body	N Shore	0.80	0.83	0.83	3.15E−05	7.58E−06	1.02E−07	0.05
cg00004209			N Shelf	0.74	0.78	0.78	0.000201	5.49E−05	6.12E−07	0.07
cg26213873	CTTNBP2NL	5'UTR	Island	0.10	0.12	0.12	1.03E−05	2.32E−05	8.21E−07	0.07
cg02678356	ZXDA	1stExon	S Shore	0.22	0.28	0.29	0.000266	1.27E−05	8.49E−07	0.07
cg09978321	CEBPG	TSS200	Island	0.02	0.03	0.03	0.000122	5.81E−06	9.01E−07	0.07
cg03033398	ZNF746	Body	S Shelf	0.83	0.85	0.85	5.20E−05	3.61E−05	9.10E−07	0.07
cg27044202	TRIM66	TSS1500		0.86	0.88	0.88	6.30E−05	4.81E−05	1.73E−06	0.11
cg21050392	HYLS1	5'UTR	S Shore	0.06	0.07	0.07	0.000135	1.17E−05	1.76E−06	0.11
cg07832337	ATP2C2	5'UTR	Island	0.11	0.06	0.06	4.03E−05	0.000115	2.07E−06	0.11
cg16131534	TBC1D22A	TSS1500	Island	0.05	0.06	0.06	0.000473	3.18E−06	2.64E−06	0.13
cg03589311	VPS52	Body	S Shelf	0.88	0.89	0.89	0.000831	1.30E−05	3.52E−06	0.13
cg23246509	TMEM109	3'UTR	N Shelf	0.83	0.85	0.85	7.20E−05	0.000151	3.72E−06	0.13
cg02800384	BANP	Body	Island	0.88	0.90	0.89	0.002061	5.30E−05	5.15E−06	0.13
cg17714794	BNIP1	1stExon		0.06	0.07	0.07	0.000365	5.21E−05	5.19E−06	0.13
cg05497240	C5orf4	3'UTR		0.83	0.85	0.85	8.89E−06	0.000503	5.32E−06	0.13
cg23363818	ZNF433	Body	N Shelf	0.73	0.76	0.76	0.000453	5.95E−05	5.90E−06	0.13
cg07086112	RHOBTB2	Body	N Shelf	0.77	0.79	0.79	0.000840	6.17E−05	6.07E−06	0.13
cg20258580	IGF1R	Body	Island	0.85	0.86	0.87	0.000107	0.000146	6.19E−06	0.13
cg00164894	USP24	Body		0.79	0.81	0.81	0.000189	0.000253	6.30E−06	0.13
cg24885794	SGCE	TSS1500	Island	0.39	0.42	0.42	0.000854	4.57E−05	6.32E−06	0.13
cg03279631				0.82	0.85	0.86	0.000875	0.000682	6.38E−06	0.13
cg26248486	BBS10	TSS200S	Shore	0.07	0.07	0.07	1.00E−05	0.000338	6.79E−06	0.13
cg12740512	C20orf94	5'UTR	S Shore	0.05	0.05	0.05	0.000414	1.66E−05	6.89E−06	0.13
cg13563193	PDCD5	Body	Island	0.04	0.05	0.05	0.000878	2.97E−05	7.01E−06	0.13
cg01534273	FAM108A1	Body	Island	0.55	0.60	0.61	0.000266	0.000110	7.08E−06	0.13
cg18634443	TBPL1	3'UTR		0.63	0.66	0.67	1.54E−05	0.000345	7.27E−06	0.13
cg03721017	ELL	Body	S Shore	0.85	0.87	0.87	0.000182	0.000649	7.49E−06	0.13
cg03936229	MSI2	Body		0.81	0.82	0.83	0.000153	0.000244	7.59E−06	0.13

TABLE 17-continued

The Top 30 Most Significantly Associated Probes in the Abstinent vs Pooled Moderate and Heavy Drinkers Analysis

				Average Methylation						BH Corrected
Probe ID	GENE	Placement	Island Status	Abst	Mod	Heavy	H vs Abs	Mod vs Abs	H&Mod vs Abs	Value
cg07574621	XPC	TSS200	Island	0.02	0.03	0.03	9.96E−05	0.000307	7.78E−06	0.13
cg09008753	SMAP1	Body	S Shore	0.06	0.06	0.06	0.000709	2.72E−05	9.12E−06	0.13

* Nominal P-value before Benjamini-Hochberg Step Up correction,

Abbreviations: Abs = abstinent, Mod = moderate, H = heavy, S = south, and N = north.

All methylation values are average beta values.

TABLE 18

Relative enrichment of CpG values with respect to Island Status and Extent of Alcohol Use

Location	All Probes		Mild User		Moderate User		Heavy User	
Island	150254	30.94%	47	36.7%	203	37.9%	894	52.3%
S_Shore	49197	10.13%	6	4.7%	26	4.9%	48	2.8%
N_Shore	62870	12.95%	17	13.3%	67	12.5%	200	11.7%
N_Shelf	24844	5.12%	7	5.5%	21	3.9%	36	2.1%
S_Shelf	22300	4.59%	21	16.4%	73	13.6%	186	10.9%
No Annotation	176112	36.27%	30	23.4%	146	27.2%	347	20.3%
	485577		128		536		1711	

TABLE 19

The Top 30 Most Significantly Associated 11 Probe Regions.

				Average Methylation			
Probe ID	GENE	Placement	Island Status	Abs	Heavy	P value*	Step Up P-value
cg24338351	BLCAP	5'UTR	Island	0.69	0.76	1.86E−10	4.75E−05
cg24675557	BLCAP	5'UTR	Island	0.68	0.75	2.36E−10	4.75E−05
cg01466133	BLCAP	5'UTR	Island	0.71	0.77	3.86E−10	4.75E−05
cg20479660	BLCAP	5'UTR	Island	0.68	0.74	3.91E−10	4.75E−05
cg07557337	RAB1B	TSS200	Island	0.07	0.07	1.29E−08	0.0011
cg26522319	C1orf103	1stExon	Island	0.05	0.06	1.53E−08	0.0011
cg02898883	RAB1B	TSS1500	Island	0.07	0.07	1.69E−08	0.0011
cg03436478	SGCE, PEG10		S_Shore	0.51	0.52	2.28E−08	0.0013
cg09337653	SIN3A, SIN3A		Island	0.08	0.08	3.29E−08	0.0016
cg27535677			N_Shore	0.70	0.75	3.30E−08	0.0016
cg20041873	SGCE, PEG10		S_Shore	0.51	0.53	3.86E−08	0.0017
cg07156273	BLCAP	5'UTR	Island	0.71	0.77	4.87E−08	0.0019
cg24141738	ERMAP, CCDC23		Island	0.04	0.05	5.25E−08	0.0019
cg15473473	BLCAP	5'UTR	Island	0.71	0.77	6.14E−08	0.0020
cg04303139	SGCE, PEG10		S_Shore	0.51	0.52	6.34E−08	0.0020
cg01758634	SIN3A	TSS1500	Island	0.07	0.08	6.87E−08	0.0020
cg05509218	SGCE, PEG10		S_Shore	0.49	0.51	8.89E−08	0.0025
cg22421148	BLCAP	5'UTR	Island	0.66	0.73	9.73E−08	0.0026
cg03759229			N_Shore	0.68	0.73	1.14E−07	0.0029
cg22893248	ACTR3C	1stExon	Island	0.07	0.10	1.25E−07	0.0030
cg20631204	ZNF562	TSS200	Island	0.11	0.13	1.34E−07	0.0031
cg02639123	ACTR3C, LRRC61		S_Shore	0.21	0.23	1.47E−07	0.0031
cg01959416	ACTR3C, LRRC61		S_Shore	0.07	0.10	1.47E−07	0.0031
cg12862537	BLCAP	5'UTR	Island	0.76	0.82	1.63E−07	0.0031
cg22497095			N_Shore	0.70	0.75	1.64E−07	0.0031
cg21516287	ACTR3C, LRRC61		S_Shore	0.14	0.16	1.68E−07	0.0031
cg26544607	ARRDC3		S_Shelf	0.07	0.07	1.88E−07	0.0032
cg04558861	LIN37	TSS200	N_Shore	0.12	0.13	1.89E−07	0.0032
cg22510412	BLCAP	5'UTR	Island	0.64	0.70	2.00E−07	0.0032
cg19998456	GGA1	TSS200	Island	0.05	0.06	2.03E−07	0.0032

*Nominal P-value before Benjamini-Hochberg Step Up correction.

Abbreviations: Abs = abstinent, S = south, and N = north.

All methylation values are average beta values,

TABLE 20

The Top 30 Most Differentially Regulated Gene Ontology Pathways

		Genes		Log^{10}	
GO Category	Category Name	Total	Changed	P-Value	FDR
GO:0005622	intracellular	1123	1951	21.91	0
GO:0044424	intracellular part	10940	925	−19.83	0
GO:0044237	cellular met. process	7595	690	−18.24	0
GO:0006139	nucleobase nucleoside nucleotide and nucleic acid met. process	4399	440	−16.19	0
GO:0044260	cellular macromolecule met. process	5765	543	−15.58	0
GO:0034641	cellular nitrogen compound met. process	4797	467	−15.22	0
GO:0005634	nucleus	5248	500	−14.80	0
GO:0043231	intracellular membrane bounded organelle	8408	732	−14.74	0
GO:0043227	membrane bounded organelle	8414	732	−14.66	0
GO:0043229	intracellular organelle	9334	795	−14.43	0
GO:0090304	nucleic acid met. process	3811	385	−14.34	0
GO:0043226	organelle	9349	795	−14.21	0
GO:0044238	primary met. process	7696	678	−14.06	0
GO:0006807	nitrogen compound met. process	4893	468	−13.82	0
GO:0008152	met. process	8521	734	−13.53	0
GO:0044428	nuclear part	2407	262	−12.72	0
GO:0043170	macromolecule met. process	6318	569	−12.42	0
GO:0005488	binding	11509	925	−11.14	0
GO:0044446	intracellular organelle	5465	498	−11.10	0
GO:0044422	organelle part	5532	501	−10.73	0
GO:0031323	regulation of cellular met. process	3834	368	−10.37	0
GO:0050794	regulation of cellular process	6319	557	−10.23	0
GO:0080090	regulation of primary met. process	3632	350	−9.99	0
GO:0051325	interphase	329	58	−9.96	0
GO:0005654	nucleoplasm	1204	145	−9.84	0
GO:0034645	cellular macromolecule biosynthetic process	3598	346	−9.74	0
GO:0009059	macromolecule biosynthetic process	3664	350	−9.51	0
GO:0051171	regulation of nitrogen compound me process	3153	309	−9.47	0
GO:0007049	cell cycle	1142	138	−9.47	0
GO:0005515	protein binding	6815	589	−9.47	0

metabolic = met.,
FDR = false discovery rate.

Example 8

Coordinated Changes in AHRR Methylation in Lymphoblasts and Pulmonary Macrophages from Smokers

[0213] Despite extensive preventative and treatment interventions, approximately 19% of American adults smoke on a daily basis (Center for Disease Control (CDC) 2011). This is a substantial problem because smoking is the leading preventable cause of premature morbidity and mortality. Smoking causes approximately 450,000 premature deaths annually through its effects on the incidence of cancer, heart disease and chronic obstructive pulmonary disease (CDC 2005). National data indicate that while both prevalence of smoking and mortality from lung cancer have significantly decreased for men between 1975 and 2007, these rates did not decrease for any racial or ethnic group or for women. In addition, projections suggest that because women who were born around 1960 have higher prevalence of smoking and morbidity than other cohorts, this gender disparity may increase.

[0214] Many of the effects of smoking on the lung are thought to result from the direct effects of cigarette smoke on pulmonary epithelium and alveolar macrophages. However, the exact mechanism(s) through which smoking increases the risk for disease in non-pulmonary tissues such as blood and brain are unclear. Recently, sets of convergent findings have suggested that a portion of that vulnerability may be driven by differential DNA methylation acquired by smoking.

[0215] Altered DNA methylation that results from genetic lesions present at conception has long been established as a cause of disorders affecting early development of disease in the soma and the CNS. With respect to non-CNS disease, altered imprinting that usually results from maternal monosomy at 15QQ causes Prader-Willi syndrome. With respect to the CNS disease, almost all cases of Rett Syndrome result from mutations in MECBP2 which exert their effects by altering DNA methylation. Guided by clues such as the observations that addition of folate, a methyl donor, to the diets of pregnant women, markedly decreases the frequency of neural tube defects, the field has embraced the concept that alterations in DNA methylation may be associated with acquired early onset developmental disorders as well. However, whether environmentally acquired alterations could increase likelihood of disease in adults has been an open question. A number of single gene and genome wide studies provide evidence that altered DNA methylation is associated with smoking and may be a cause of smoking associated illness. In particular, using both genome wide and single gene approaches, we have demonstrated that altered DNA methylation is associated with smoking (see above Examples). However, these studies have been hindered by low coverage of the total number of genes and CpG residues in the human genome and discrepancies as to the appropriateness of certain forms of biomaterials for studies of epigenetic phenomena.

[0216] In this communication, we report our results with respect to smoking status on genome wide methylation and focal gene expression using two independent sets of biomaterials: 1) lymphoblast DNA and RNA derived from 119 female subjects from the Iowa Adoption Studies (IAS) and 2) alveolar macrophage DNA from cells isolated from the lungs of 10 smokers and 9 non-smokers.

Methods

[0217] Human Subjects. The first set of biomaterials was obtained from subjects participating in the Iowa Adoptions Studies (IAS) (Yates et al., 1998, The Iowa Adoption Studies Methods and Results. In: LaBuda et al., Ed's, *On the Way to Individuality: Methodological Issues in Behavioral Genetics*, Hauppauge N.Y., Nova Science Publishers, pp 95-125). In brief, the IAS is a case and control adoption study of the role of genetic, environmental and gene-environment interactions in the etiology of common behavioral illness. The clinical material used in the current study is derived from interviews with the Semi-Structured Interview for the Assessment of the Genetics of Alcoholism, Version II (Bucholz et al., 1994, A new, semi-structured psychiatric interview for use in genetic linkage studies: a report on the reliability of the SSAGA, *J. Stud. Alcohol,* 55:149-58), during each of the last two waves of the IAS study (1999-2004 and 2005-2009). The biological material used in this study, lymphoblast cell lines, was

derived by Epstein Barr virus mediated transformation (Caputo et al., 1991, An Effective Method for Establishing Human B Lymphoblastic Cell Lines Using Epstein Barr Virus, *J. Tiss. Cult. Meth.,* 13:39-44) of lymphocytes obtained from blood donated by 165 female subjects during the last wave of the study.

[0218] The second set of biomaterials for the current study was alveolar macrophages obtained by bronchoalveolar lavage. Subjects were recruited from the community via advertisements and word-of-mouth. In order to be included, case (smoking) subjects had to be actively smoking with at least 10 pack year history of smoking. To be included as a control, the subject had to deny ever smoking cigarettes. Subjects were excluded if they had any significant co-morbid conditions such as pregnancy, or if a baseline spirometry revealed the Forced Expiratory Volume in the first second (FEV1) was less than 60% of predicted. All of these procedures and protocols were approved by the University of Iowa Institutional Review Board.

[0219] Bronchoalveolar Lavage. To obtain human alveolar macrophages, a bronchoalveolar lavage was performed. After informed consent was obtained, subjects underwent standard flexible bronchoscopy. After the application of local anesthesia, bronchoalveolar lavage was performed by instilling 20 ml of normal saline into a tertiary bronchus up to five times in three different lung segments. The first collection out of five was discarded for possible contamination from upper airway secretions or by lidocaine, which is used to locally anesthetize the subject during the procedure. The remaining lavage was transported to the laboratory where fluid was filtered through sterile gauze and centrifuged at 200×g for 5 min to pellet cellular material. The resulting pellet was suspended in phosphate buffered saline and centrifuged at 16,000×g for one minute. The macrophages were suspended in medium, labeled with Wright stain and microscopically examined to ensure that greater than 95% of the cells were macrophages.

[0220] DNA and RNA Isolation. The lymphoblast DNA and RNA used in this study was prepared from growth-entrained cell lines according to our standard procedures (Philibert et al., 2008, MAOA methylation is associated with nicotine and alcohol dependence in women, *Am. J. Med. Genet. B. Neuropsychiatr. Genet.,* 147B:565-70). In brief, on the day before DNA preparation, one-half of the cell media for each culture flask was exchanged. Twenty four hours later, DNA was prepared from the cell lines using cold protein precipitation. Simultaneously, RNA was purified from independent aliquots of the same culture using RNA Midi kits (Invitrogen, USA) according the instructions of the manufacturer. After quantification and purity assessment using a Nanodrop (Thermo Scientific, USA) spectrophotometer, DNA was stored at −20° C. and RNA was stored at −80° C. until use.

[0221] DNA and RNA were isolated from alveolar macrophages using the Qiagen DNAeasy™ kit (Qiagen, Valencia, Calif.) and MirVana (Applied Biosystems, Austin, Tex.) reagents according to manufacturer's instructions. Quality assessment was by Nanodrop and Experion (Bio-Rad Experion Automated Electrophoresis Station). After preparation, DNA was stored at −20° C. and RNA was stored at −80° C. until use.

[0222] DNA Methylation. Genome wide DNA methylation of the DNA was assessed using the Illumina HumanMethylation450 BeadChip under contract by the University of Minnesota Genome Center using the protocol specified by the manufacturer and the contractor. The resulting microarray data were inspected for complete bisulfite conversion of the DNA, and average beta values (i.e. average methylation) for each CpG residue were determined using the GenomeStudio V2009.2; Methylation module Version 1.5.5, version 3.2 (Illumina, San Diego). The resulting beta values were exported into Microsoft Excel and JMP (SAS Institute, USA) for data analysis. The HumanMethylation450 BeadChip contains 485,577 probes that recognize at least 20216 unique features (i.e. potential transcripts). With respect to this sample, >99.76% of the 485,577 probes yielded statistically reliable data.

[0223] Data Analysis. After logarithmic conversion, data were inspected for outliers or confounding by plate or chip variables, then the initial data analyses were conducted using genome wide t-tests. Subsequently, beta values for each of the probes were aligned according to their physical location and the data re-analyzed using paired t-tests over a 11-probe sliding window in order to more adroitly capture methylation signatures over larger regions (Dindot et al., 2009, Epigenetic profiling at mouse imprinted gene clusters reveals novel epigenetic and genetic features at differentially methylated regions, *Genome Res.,* 19:1374-83; Farthing et al., 2008, Global Mapping of DNA Methylation in Mouse Promoters Reveals Epigenetic Reprogramming of Pluripotency Genes, *PLoS Genet.,* 4:e100-16). All genome wide comparisons were corrected for multiple comparisons using the method of Benjamini and Hochberg (1995, Controlling the false discovery rate: A practical and powerful approach to multiple testing, *J. Royal Statistical Soc., Series B, Methodological,* 57:289-300). For select loci, data were analyzed with respect to alcohol use status using ANOVA.

[0224] Pathway analysis of differentially methylated genes was conducted using GoMiner™ using default settings (0.05 settings for reports and all gene ontology as the root category setting) using the gene set specified in the text as the "changed" gene set (Zeeberg et al., 2003, GoMiner: a resource for biological interpretation of genomic and proteomic data, *Genome Biol.,* 4:1-8). All values reported include nominal and FDR corrected values.

[0225] Specific qRT-PCR Analysis of AHRR. The relative expression of AHRR was determined using primer probe sets from ABI, a Fluidigm BioMark™ System and proprietary BioMark Real-Time Analysis software according to manufacturer's guidelines. Briefly, first, RNA was converted to cDNA using an ABI cDNA archiving kit according to manufacturer's suggestions. Then after a brief pre-amplification step, each cDNA sample was amplified in quadruplicate with using primer probes for AHRR (Hs01005075) and five housekeeping genes (CALR, RPL7A, PRS19, RPS20 and UBC) obtained from Applied Biosystems (Foster City, USA). The Ct counts exported to the database, normalized using the geometric mean of five housekeeping genes, then converted to Z scores for statistical analysis.

Results

[0226] Iowa Adoption Study Cohort. The demographic and clinical characteristics of the 165 female subjects whose genome wide methylation status was assessed are shown in Table 21. Overall, the subjects were largely white and tended to be in their mid-to-late 40s. Consistent with enrichment of the sample for the diathesis of substance use, the majority of the subjects in the study reported daily smoking at some period of their lives (85 of 165). However, many of these individuals (n=46) have quit smoking or were not smoking

every day at the time of phlebotomy leaving only 39 subjects reporting daily smoking (i.e. seven days per week every week) at the time of phlebotomy. Because our prior studies have indicated that they methylation signature of those subjects who had recently quit smoking is highly variable, those 46 individuals were excluded from further study (Philibert et al., 2010, The effect of smoking on MAOA promoter methylation in DNA prepared from lymphoblasts and whole blood, *Am. J. Med. Genet. B. Neuropsychiatr. Genet.*, 153B:619-28). The number of cigarettes smoked daily by the 39 subjects who smoked daily varied from 4 to 40 with the average number of cigarettes consumed daily being 19 cigarettes or about a pack per day for greater than 20 years. Cigarette smoking tended to be the only form of nicotine use currently being manifested by these 39 subjects with none of the subjects reporting the concomitant use of cigars, chew or other forms of nicotine usage in 2 weeks prior to assessment. There were no significant differences between the three groups (current smokers, never smokers, non-daily smokers/quitters) with respect to alcohol use in the past six months or age.

[0227] We contrasted the methylation values for the 39 smokers (average beta value 0.443) with the values for the 80 non-smokers (average beta value 0.446) using single point genome wide t-tests. The results of those analyses are shown in Table 22. As the table indicates, only one probe, cg14817490, which maps to intron 3 of the of the aryl hydrocarbon receptor repressor (AHRR), survived genome wide Benjamini-Hochberg correction for multiple comparisons. However, it is interesting to note that 3 other probes from AHRR, cg05575921, cg14454127, and cg03991871, were ranked among the top 13 probes and that none of them were from the rather small promoter associated CpG island. Instead, all 4 of the top AHRR probes target the gene body which contains three (>100 CpG residues) large CpG island according the UCSC genome browser. Finally, we note that cg03636183, a probe that was reported to be significantly associated with smoking status in lymphocyte DNA, was also nominally associated ($p<0.003$; rank 802nd of 485577 probes; smoker average 0.67; non-smoker average 0.74) with smoking status in the current study (Breitling et al., 2011, Tobacco-Smoking-Related Differential DNA Methylation: 27K Discovery and Replication, *Am. J. Human Genet.*, 88:450-7).

[0228] One possible concern is that some of the differential methylation signature could be secondary to alcohol use. Therefore, even though there were no significant differences between the rate of drinking for smoker and non-smoker groups, we analyzed the data for alcohol-related changes. The relationship of methylation to alcohol intake over the past 6 months to the methylation at loci controlling for alcohol use status was examined. Only two of the top 30 probes, cg07812589 and cg17231418, were even nominally related to amount of alcohol intake in the past 6 months, both at a p-value of $0.04<x<0.05$. Hence, there does not appear to be any effect of alcohol intake on the methylation status at the most differentially methylated loci.

[0229] Next, as part of our analyses, we conducted a sliding window analysis using an 11-probe window and the same groups of case and control subjects. Table 23 describes the result of those analyses. The addition of the methylation data immediately flanking each probe increased the overall significance of the findings with 36 comparisons surviving genome wide correction. Not surprisingly, many of the top thirty probes from the analysis tended to lie immediately adjacent to one another. Interestingly, despite the strength of four AHRR probes in the single probe analyses, the gene region containing these probes, which is interrogated by 149 separate markers, was not included in this list of top regions. Inspection of this locus shows that differential methylation was largely confined to the 2 or 3 probe windows surrounding each of these residues with each of these areas being several thousand base pairs apart (Appendix A).

[0230] Using GoMiner™, we conducted gene pathway analyses using the information from the 273 probes that were nominally differentially methylated at the $p<0.001$ level. Table 24 shows the top 30 most differentially methylated pathways. Overall, only one pathway, protein kinase C (PKC) activity, survived false discovery rate (FDR) correction at the $p<0.05$ level. However, a recurrent theme of differential methylation in gene pathways affecting ion transport was found in many of the other less significant top thirty pathways.

[0231] Human Alveolar Macrophage Data. Because some may have concerns about the reliability of lymphoblast ability to model the changes found in their cognate lymphocytes and other primary cell types, we repeated these same case and control analyses using DNA from pulmonary alveolar macrophages again using a case and control paradigm. The case macrophages were isolated from the lungs of 10 smokers with at least a 10 year history of ≧1 ppd smoking (6 male and 3 female) while the control macrophage biomaterial set was isolated from 9 non-smokers (6 male and 4 female). Although these two groups were roughly matched for ethnicity (smokers: 8 White, 2 African Americans; non-smokers: 9 White), the control group was significantly younger than the smoking group (smokers 31±3 yrs, non-smokers 40±4 yrs, $p<0.01$).

[0232] The results of the genome wide single probe contrasts are illustrated in Table 25. Overall, the effects of smoking were much more profound with 1381 probes surviving correction for genome wide comparison at a $p<0.05$ level. Of considerable interest given recent data suggesting a prominent role for AHRR in carcinogenesis, 8 probes from AHRR, including the 3rd ranked probe, cg25648203, were significantly associated after correction for genome wide comparisons. But of the top 4 AHRR probes from the lymphoblast analyses, only cg05575921 was significantly associated after Bonferroni correction.

[0233] We next repeated the sliding window analyses for the macrophage data using the same method delineated above. Once again, the results (see Table 26) were more robust than those for the lymphoblast data with 40 eleven probe regions being significantly associated after correction for multiple comparisons. Although many highly interesting genes were once again implicated in this analysis, AHRR was once again notable with the 28th ranked 11 probe region being found in the body of the AHRR.

[0234] As a last part of our set of analyses with respect to the macrophage methylation data, we repeated the GoMiner pathway analyses using the list of 1381 probes which were significantly associated in the above analyses as our changed gene set. Table 27 shows those results of those analyses. In brief, pathways involved with wound healing, inflammation and G-protein/ras signaling were particularly prominent.

[0235] Comparison of Lymphoblast and Macrophage Data. In both the macrophage and lymphoblast analyses, probes from AHRR were repeatedly associated with smoking status. Therefore, we compared the methylation signatures from these two biomaterials with respect to smoking status. Appen-

dix A details the average methylation and single point analyses for each of the 146 probes for the gene for each biomaterial. In brief, 14 probes in the lymphoblast analyses and 40 of the probes in the macrophage analyses were associated with smoking status at a $p<0.05$ with 8 of the 14 probes in the lymphoblast analyses also being nominally significantly associated with smoking status in the macrophages with the direction of methylation being consistent at each probe (greater methylation in smokers). The overall methylation signature between the control lymphoblasts and macrophages at AHRR was highly correlated ($r=0.95$). FIG. **9** illustrates the relationship between the differential methylation at each of the 146 residues listed in Appendix A for the lymphoblast and macrophage DNA samples. As FIG. **9** shows, the differential methylation signature was also highly correlated across the gene with over 20% of the differential methylation signature that was associated with smoking status being shared between the two DNA sources ($r=0.45$; $p<0.001$).

[0236] An advantage of lymphoblasts is the ability to easily create high-quality RNA for gene expression studies. Therefore, to determine whether this differential methylation had functional consequences on lymphoblast gene expression, we then analyzed the relationship between AHRR gene expression and methylation status at cg05575921, the AHRR probe with the most consistent associations in the two analyses, using RNA prepared from the case and control samples. Interestingly, increasing methylation at this probe was associated with decreasing lymphoblast AHRR gene expression ($p<0.03$, $n=108$), which suggests that the CpG residues in this region may have a functional in vivo role in regulating gene expression at this locus.

Discussion

[0237] In summary, we report that cigarette smoking is associated with significant changes in genome wide methylation, and in particular, AHRR methylation, in DNA derived from pulmonary alveolar macrophages and lymphoblasts. Strengths of this manuscript include confirmation of the findings from lymphoblast DNA, which are immortalized lymphocytes, with data from primary tissue from the lungs of smokers and the presentation of evidence that these changes at AHRR may be functional. Possible limitations include the relative poor matching of the subjects who contributed lymphoblast and pulmonary macrophage DNA.

[0238] The most significant and consistent finding in the current study is with respect to AHRR locus. AHHR is a feedback inhibition modulator of AHR that exerts its effects by competing with AHR for binding with its cognate nuclear receptor dimer partner (AHR nuclear translocator) or at xenobiotic response elements in AHR regulated genes. This feedback modulation plays a pivotal role in AHR regulation and may be critical in moderating AHR role in oncogenesis and altered immune function. Our finding of smoking associated methylation at AHRR is highly plausible for several reasons. First and foremost, smoking is the leading preventable cause of cancer. Hence, this association may explain part of the connection. Second, the direction of the differential methylation was consistent among the 8 AHRR probes with nominal significance in both lymphoblast and macrophage comparisons with a high degree of shared smoking associated differential methylation. Third, AHRR was the only gene locus that had significant localizations in both studies after correction for multiple comparisons. Fourth, previous studies have shown that smoking induces production of the aryl hydrocarbon receptor (AHR), a process which is thought to be critical for certain forms of smoking related forms of carcinogenesis. Assuming that the decreased methylation at AHRR seen in smokers in the current study may result from a feedback mechanism associated with smoking induction of AHR transcription, the current findings are very consistent with previous findings and suggest potential avenues for addressing AHR mediated neoplastic transformation. Unfortunately, even given the promising gene expression findings, rigorous testing of this hypothesis may be difficult because review of the Ensembl and UCSC genome browser databases demonstrates the presence of three large CpG islands that are interspersed throughout the gene and at least 11 AHRR transcripts, each of which codes for a differently sized protein that may have unique competitive properties with respect to AHR. Hence, while the current findings are encouraging, a more definitive understanding of relationship between AHHR methylation and both AHRR gene expression and AHR function may require more complex and detailed examination of this region.

[0239] The pathway analyses of the macrophage data were illuminating and consistent with our understanding of the effects of smoking. The macrophage data was characterized by changes in inflammation, wound healing and Ras/G-protein signaling pathways. The repeated finding of altered methylation in Ras/G-protein signaling pathways seems logical since activation of these proteins are thought to be part of the oncogenic process for many types of cancers. Similarly, the recurrent identification of wound healing and inflammatory pathways seems logical since smoking is the leading cause of Chronic Obstructive Pulmonary Disease (COPD), a syndrome in which the vast morbidity of the pathology is secondary to inflammatory moderated remodeling of the lung epithelium. In contrast, the results of the lymphoblast analyses were less robust with only two pathways, related to peptidyl-threonine modification (PKC), surviving FDR correction. However, it is important to note that, while both pathways are closely related, with the basis of their significance in our analyses relying on the same five probes, the omission of one probe from either of these comparisons would result in nonsignificant findings.

[0240] The comparative weakness of the methylation findings in lymphoblasts as compared to macrophages highlight the importance of incorporating studies of primary tissues directly exposed to the substance in question. Overall, the smoking associated differential methylation was markedly more pronounced in the alveolar macrophage DNA than in the lymphoblast DNA. This is probably because circulating lymphocytes are less exposed to the direct effects of smoke than the macrophages resident in the lung. However, it is possible that our conversion of these same lymphocytes into the transformed lymphoblast cell lines may further weaken the smoking induced signal. The latter possibility needs to be considered because although lymphoblast cell lines are excellent models of the lymphocytes from which they are derived, lymphoblast lines are vulnerable to clonal selection artifacts and there are well documented differences between lymphocyte and lymphoblast gene expression. The fact that the lymphoblasts by definition proliferate in non-smoking conditions, may impact the data. If the smoking methylation changes are dependent on continued in vivo smoke exposure, then replication in culture may mute the findings. This supports the importance of examining primary cells along with lymphoblasts.

[0241] It should be recognized that most investigators use Ficoll separated mononuclear cell pellets rather than purified lymphocytes. Since these "lymphocyte pellets" contain a variety of cell types including B-lymphocytes, T-lymphocytes, monocytes and Natural Killer T-cell, it may well be that use of this heterogeneous cell mix may have obscured other potential findings which may explain why only one differentially methylated probe was previously identified despite using a similar number of subjects.

[0242] Beyond the relative merits of lymphocyte and lymphoblast preparations, the current findings suggest that the lymphoblast lines paired with primary pulmonary macrophages will be useful in other investigations of the epigenetics of smoking because: 1) smoking has a broad effect on tissues throughout the body including the blood, and 2) integration of histone modification and gene expression status with DNA methylation status will require large numbers of cells. Some types of histone modification examinations necessitate relatively larger amounts of fresh cellular material. This suggests the utility of lymphoblasts in histone modification studies. A clear picture of lymphoblast gene expression and DNA methylation data relative to a primary smoking-relevant cell (alveolar macrophages) data will be needed for these potential future studies. In this respect, our convergent finding in lymphocytes and macrophages with respect to AHRR are reassuring.

[0243] One potential direction for future work is the determination of the specific AHRR transcripts that are differentially affected by differential methylation. The Taqman™ gene expression probe for AHRR used in this study (Hs01005075) recognizes the exon 3-exon 4 boundary that is included in most splice variants. However, given the numerous splice variants produced by this gene, the epigenetic complexity of the gene (e.g. three large CpG islands not associated with the promoter), and its putative role in oncogenesis, future studies that examine specific splice variants altered by smoking is warranted.

[0244] The relationship of gene methylation to histone code modification should also be explored. In particular, the relationship of H3K4 and H3K27 methylation and H3K27 acetylation to AHRR gene expression should be examined because of the strong relationship of these modifications to gene expression. Though DNA methylation is thought to have a weaker relationship to gene expression, if we can establish a stronger understanding of the histone-DNA modification relationship on a genome wide level, it well may be that we can use DNA methylation at loci such as AHRR as a proxy for histone status, and thereby gene expression status. Studies of DNA methylation are much cheaper and easier to conduct than histone modification studies. A better understanding of the relationship of peripheral blood methylation to methylation in other tissues, such as brain, may allow more informative studies of the role of DNA methylation and other forms of epigenetic changes in normal and disease related human development.

[0245] In summary, we report that cigarette smoking is associated with genome wide changes in lymphoblast and pulmonary macrophage DNA methylation, in particular at AHRR. We suggest replication and extension of the current findings and further investigations of the role of epigenetic changes in smoking altered gene expression.

TABLE 21

Clinical Characteristics of the 165 Female Iowa Adoptions Studies Probands

	Non-Smoker	Quit or Quitting	Daily
Smoker			
N	80	46	39
Age	46 ± 8	47 ± 8	43 ± 6
Ethnicity			
White	80	44	39
Other	0	4	0
Alcohol in Past 6 months			
Yes	58	35	29
No	22	11	7
Daily Cigarette Usage			19 ± 9

TABLE 22

The Top 30 Most Significantly Differentially Methylated Probes in Lymphoblast DNA

Probe ID	GENE	Placement	Island Status	N-Smoker Avg	Smoker AVG	T-test	Corrected P-value
cg14817490	AHRR	Body		0.24	0.12	2.71E−08	0.02
cg05575921	AHRR	Body	N Shore	0.85	0.70	1.34E−06	0.29
cg07313705		S Shelf		0.07	0.10	1.78E−06	0.29
cg14454127	AHRR	Body		0.44	0.31	2.72E−06	0.34
cg02486161	NOD2	3'UTR		0.70	0.59	2.53E−05	0.99
cg14983684	RAD51L1	Body		0.75	0.71	2.58E−05	0.99
cg23939642	SLC38A10	Body		0.50	0.33	2.66E−05	0.99
cg25325005	PLEC1	Body	N Shelf	0.63	0.41	2.96E−05	0.99
cg23335946	C1orf25	1stExon	Island	0.08	0.09	3.14E−05	0.99
cg20776920	UNC5D	TSS1500	N Shore	0.87	0.83	3.21E−05	0.99
cg26812418	CPE	TSS200	Island	0.05	0.07	4.09E−05	0.99
cg07812589				0.26	0.20	4.59E−05	0.99
cg03991871	AHRR	Body	N Shore	0.78	0.67	4.97E−05	0.99
cg27545205		Island		0.02	0.02	5.26E−05	0.99
cg10951975	TRPM4	Body	Island	0.35	0.22	5.49E−05	0.99
cg20370184	SLC44A4	Body		0.27	0.12	5.64E−05	0.99
cg07999887	CPNE3	5'UTR	Island	0.02	0.02	5.92E−05	0.99
cg08644463	GNAI3	Body		0.87	0.83	6.83E−05	0.99
cg04366249	SGCE	1stExon	Island	0.05	0.07	7.34E−05	0.99
cg12741529	C3orf17	Body		0.87	0.85	7.75E−05	0.99
cg08940570	LOXL3	5'UTR	N Shore	0.80	0.66	9.09E−05	0.99

TABLE 22-continued

The Top 30 Most Significantly Differentially Methylated Probes in Lymphoblast DNA

Probe ID	GENE	Placement	Island Status	N-Smoker Avg	Smoker AVG	T-test	Corrected P-value
cg23754924	RGMA	Body	Island	0.10	0.13	9.56E−05	0.99
cg24547565	RUSC1	TSS1500	N Shore	0.51	0.62	9.85E−05	0.99
cg17093877	MGC16275	Body	N Shelf	0.57	0.43	0.00010	0.99
cg21545248	HMGXB3	Body		0.77	0.71	0.00011	0.99
cg22012583	LASS2	TSS1500	Island	0.37	0.25	0.00011	0.99
cg17231418	ESX1	Body	Island	0.26	0.39	0.00011	0.99
cg12668122	TMEM108	Body		0.40	0.31	0.00012	0.99
cg19776793	SLC38A10	Body		0.43	0.25	0.00013	0.99
cg02724404	LYSMD4	TSS1500	S Shore	0.88	0.84	0.00013	0.99

All average methylation values are non-log transformed beta-values. Island status refers to the position of the probe relative to the island. Classes include: 1) Island, 2) N (north) shore, 3) S (south) shore, 4) N (north shelf), 5) S (south) shelf and 6) blank denoting that the probe does not map to an island.

TABLE 23

The Top 30 Most Significantly Differentially Methylated Regions in Lymphoblast DNA

				Average Methylation			
Probe ID	GENE	Placement	Island Status	N-Smoker	Smoker	P value*	Corrected P-value
cg13581859	HLA-DPB1	Body	Island	0.66	0.79	2.31E−09	0.002
cg25511667	HLA-DPB1	Body	Island	0.69	0.85	7.34E−09	0.002
cg14801692	HLA-DPB1	Body	Island	0.62	0.70	1.40E−08	0.003
cg03636880	HLA-DPB1	Body	Island	0.64	0.77	1.81E−08	0.003
cg01132696	HLA-DPB1	Body	Island	0.64	0.81	2.30E−08	0.003
cg10850215	HLA-DPB1	Body	Island	0.64	0.76	3.07E−08	0.003
cg02692313	HLA-DPB1	Body	Island	0.66	0.83	4.14E−08	0.003
cg03229061	HLA-DPB1	Body	Island	0.62	0.71	4.53E−08	0.003
cg17588455	HLA-DPB1	Body	Island	0.62	0.73	5.55E−08	0.003
cg19990651	HLA-DPB1	Body	Island	0.65	0.83	6.80E−08	0.004
cg14870156	HLA-DPB1	Body	Island	0.66	0.79	7.47E−08	0.004
cg06437840	HLA-DPB1	Body	Island	0.52	0.69	8.16E−08	0.004
cg26645432	HLA-DPB1	Body	Island	0.71	0.86	1.00E−07	0.004
cg20223237	HLA-DPB1	Body	Island	0.73	0.88	1.24E−07	0.005
cg25796439	ISM1	TSS1500	Island	0.08	0.08	1.26E−07	0.006
cg12893780	HLA-DPB1	Body	Island	0.67	0.82	1.84E−07	0.006
cg19759481	HOXA5	TSS200	Island	0.63	0.54	1.99E−07	0.007
cg04863892	HOXA5	TSS200	Island	0.68	0.60	2.53E−07	0.008
cg01992382	TNXB	Body	Island	0.42	0.47	2.74E−07	0.008
cg01370449	HOXA5	TSS200	Island	0.69	0.63	3.11E−07	0.010
cg12746059	PCDH10	TSS200	Island	0.08	0.09	3.95E−07	0.02
cg13349035	HLA-DPB1	Body	N Shore	0.68	0.80	4.72E−07	0.02
cg09549073	HOXA5	5'UTR	Island	0.68	0.60	6.91E−07	0.02
cg02916332	HOXA5	TSS1500	Island	0.64	0.58	7.89E−07	0.02
cg12128839	HOXA5	TSS200	Island	0.56	0.47	8.21E−07	0.02
cg17569124	HOXA5	TSS1500	Island	0.57	0.48	8.90E−07	0.02
cg06831576	CDH8	TSS200	Island	0.11	0.15	1.00E−06	0.02
cg04525757	FOXG1	TSS1500	N Shore	0.14	0.15	1.25E−06	0.03
cg26242583	LUZP2	TSS200	Island	0.11	0.13	1.35E−06	0.03
cg19714132	FOXG1	TSS1500	N Shore	0.19	0.21	1.58E−06	0.03

*Nominal P-value before Benjamini-Hochberg correction. Corrected value is per Benjamini-Hochberg method.

TABLE 24

The Top 30 Most Differentially Regulated Pathways in Lymphoblast DNA

		Genes		Log^{10}	
GO Category	Category Name	Total	Changed	P-Value	FDR
GO:0018107	peptidyl-threonine phosphorylation	27	5	−5.03	0.01
GO:0018210	peptidyl-threonine modification	30	5	−4.80	0.01
GO:0060914	heart formation	9	3	−4.00	0.09
GO:0009653	anatomical structure morphogenesis	1490	31	−3.53	0.15

TABLE 24-continued

The Top 30 Most Differentially Regulated Pathways in Lymphoblast DNA

		Genes		Log^{10}	
GO Category	Category Name	Total	Changed	P-Value	FDR
GO:0045121	membrane raft	160	8	−3.45	0.14
GO:0007548	sex differentiation	181	8	−3.10	0.24
GO:0005024	TGF beta receptor activity	18	3	−3.05	0.20
GO:0007530	sex determination	18	3	−3.05	0.20
GO:0003007	heart morphogenesis	104	6	−3.03	0.18

TABLE 24-continued

The Top 30 Most Differentially Regulated Pathways in Lymphoblast DNA

GO Category	Category Name	Genes Total	Genes Changed	Log^{10} P-Value	FDR
GO:0004675	transmembrane receptor protein serine threonine kinase activity	19	3	−2.97	0.19
GO:0003197	endocardial cushion development	5	2	−2.94	0.22
GO:0005026	TGF beta receptor activity type II	5	2	−2.94	0.22
GO:0060021	palate development	46	4	−2.82	0.26
GO:0051015	actin filament binding	48	4	−2.75	0.33
GO:0030501	pos. reg. of bone mineralization	23	3	−2.73	0.33
GO:0070169	pos. reg. of biomineral tissue dev.	24	3	−2.67	0.33
GO:0003128	heart field specification	7	2	−2.63	0.31
GO:0003129	heart induction	7	2	−2.63	0.31
GO:0051864	histone demethylase activity	7	2	−2.63	0.31
GO:0061311	cell surface receptor linked signaling pathway involved in heart dev.	7	2	−2.63	0.31

TABLE 24-continued

The Top 30 Most Differentially Regulated Pathways in Lymphoblast DNA

GO Category	Category Name	Genes Total	Genes Changed	Log^{10} P-Value	FDR
GO:0060389	SMAD protein phosphorylation	25	3	−2.62	0.30
GO:0001649	osteoblast differentiation	86	5	−2.62	0.29
GO:0005901	caveola	53	4	−2.59	0.29
GO:0031095	platelet tubular network membrane	8	2	−2.51	0.32
GO:0035173	histone kinase activity	8	2	−2.51	0.32
GO:0046541	saliva secretion	8	2	−2.51	0.32
GO:0045669	pos. reg. of osteoblast differentiation	28	3	−2.48	0.33
GO:0070838	divalent metal ion transport	229	8	−2.45	0.32
GO:0045778	pos. reg. of ossification	29	3	−2.43	0.33
GO:0030154	cell differentiation	2041	35	−2.43	0.32

dev. = development,

pos. reg. = positive regulation,

FDR = false discovery rate.

TABLE 25

The Top 30 Most Significantly Differentially Methylated Probes in Alveolar Macrophage DNA

Probe ID	GENE	Placement	Island Status	N-Smoker Avg	Smoker AVG	T-test	Corrected P-value
cg06961313	MR1	TSS1500		0.80	0.57	1.06E−10	5.16201E−05
cg00738897			S_Shore	0.71	0.55	1.90E−09	0.0003
cg25648203	AHRR	Body		0.38	0.72	1.97E−09	0.0003
cg00506299	RFTN1	Body		0.23	0.46	2.67E−09	0.0003
cg27229484	ZC3H12A	Body		0.26	0.53	3.34E−09	0.0003
cg05951221		Island		0.28	0.42	5.85E−09	0.0005
cg01432692				0.20	0.37	7.69E−09	0.0005
cg09374353	EHD1	3'UTR	N_Shore	0.12	0.39	1.05E−08	0.0006
cg14310198	RAPGEF1	Body		0.48	0.70	1.90E−08	0.0010
cg21566642		Island		0.33	0.56	2.12E−08	0.0010
cg17576603	DAB2	5'UTR		0.39	0.62	3.36E−08	0.0013
cg17574812	ABHD6	Body		0.26	0.49	3.55E−08	0.0013
cg06634140				0.30	0.54	3.73E−08	0.0013
cg11254522	FGR	Body		0.35	0.50	3.99E−08	0.0013
cg07457727			N_Shelf	0.22	0.60	4.02E−08	0.0013
cg13458803	CD80	5'UTR		0.36	0.16	4.86E−08	0.0014
cg01668352	SRGAP1	Body		0.32	0.62	4.97E−08	0.0014
cg04402828	KIAA1026	Body		0.47	0.35	6.69E−08	0.0018
cg07650681	LOC100132354	Body		0.66	0.40	7.19E−08	0.0018
cg13610455	LOC388796	Body		0.29	0.44	7.37E−08	0.0018
cg09127592	TRIM35	Body	N_Shelf	0.33	0.73	8.72E−08	0.0019
cg14223856				0.43	0.81	9.60E−08	0.0019
cg09006487	RYBP	3'UTR		0.35	0.50	9.69E−08	0.0019
cg02233197	TNFAIP8L3	Body	S_Shelf	0.29	0.72	9.85E−08	0.0019
cg05317600				0.34	0.65	9.86E−08	0.0019
cg25466245	SUSD4	Body		0.36	0.57	1.09E−07	0.0020
cg21418854	C1orf113	TSS1500	N_Shore	0.42	0.58	1.13E−07	0.0020
cg02341139			S_Shelf	0.34	0.60	1.17E−07	0.0020
cg18030943	LAMP3	Body	N_Shelf	0.23	0.38	1.20E−07	0.0020
cg05337681	LIPC	Body		0.23	0.47	1.25−07	0.0020

All average methylation values are non-log transformed beta-values. Island status refers to the position of the probe relative to the island. Classes include: 1) Island, 2) N (north) shore, 3) S (south) shore, 4) N (north shelf), 5) S (south) shelf and 6) blank denoting that the probe does not map to an island.

TABLE 26

The Top 30 Most Significantly Differentially Methylated Regions in Alveolar Macrophage DNA

				Average Methylation			
Probe ID	GENE	Placement	Island Status	N-Smoker	Smoker	P value*	Corrected P-value
cg07965566				0.60	0.30	2.40E−28	1.16E−22
cg14310198	RAPGEF1	Body		0.70	0.48	4.83E−26	1.17E−20
cg17574812	ABHD6	Body		0.49	0.26	9.83E−25	1.59E−19
cg01668352	SRGAP1	Body		0.62	0.32	1.76E−24	2.14E−19
cg17576603	DAB2	5'UTR		0.62	0.39	2.52E−23	2.45E−18
cg07457727				0.60	0.22	3.98E−21	3.22E−16
cg10169462				0.13	0.06	1.30E−15	9.05E−11
cg24790419	KIAA1683	TSS1500		0.62	0.39	1.54E−15	9.38E−11
cg04402828	KIAA1026;	Body		0.35	0.47	3.16E−15	1.70E−10
cg05951221				0.42	0.28	4.80E−12	2.33E−07
cg16039867	MKNK1	Body		0.58	0.78	2.50E−11	1.10E−06
cg06634140				0.54	0.30	9.81E−10	3.97E−05
cg20485084	FGR	Body		0.67	0.36	1.91E−09	7.16E−05
cg02341139				0.60	0.34	6.35E−09	0.0002
cg22019569	SMYD3	Body		0.60	0.75	1.38E−08	0.0004
cg14019523	ASB2	Body		0.37	0.24	2.11E−08	0.0006
cg13675814	CORO2A	5'UTR		0.67	0.82	6.95E−08	0.0018
cg04307274				0.57	0.31	6.96E−08	0.0018
cg15149645	NUPR1	TSS200		0.59	0.75	9.66E−08	0.0024
cg10192877	ABCG1	Body		0.61	0.74	1.20E−07	0.0028
cg21566642				0.56	0.33	1.24E−07	0.0028
cg20568305	GRAMD4	Body		0.68	0.52	2.95E−07	0.0065
cg01432692				0.37	0.20	3.35E−07	0.0070
cg24446429	MBP	Body		0.60	0.39	3.88E−07	0.0075
cg14414943	CHI3L2	Body		0.83	0.89	3.89E−07	0.0075
cg16659773				0.56	0.43	4.16E−07	0.0077
cg04135110	AHRR	Body		0.15	0.38	5.29E−07	0.0095
cg00738897				0.55	0.71	6.03E−07	0.0104
cg13458803	CD80	5'UTR		0.16	0.36	7.97E−07	0.0133
cg11691844	SYTL2	Body		0.27	0.44	8.63E−07	0.0139

*Nominal P-value before Benjamini-Hochberg correction. Corrected value is per Benjamini-Hochberg method.

TABLE 27

The Top 30 Most Differentially Regulated Gene Ontology Pathways in Macrophage DNA

		Genes		Log^{10}	
GO Category	Category Name	Total	Changed	P-Value	FDR
GO:0005737	cytoplasm	7845	429	−8.40	0
GO:0007165	signal transduction	2324	159	−7.89	0
GO:0005515	protein binding	6815	378	−7.62	0
GO:0023052	signaling	3788	233	−7.56	0
GO:0023033	signaling pathway	2813	183	−7.48	0
GO:0007264	small GTPase mediated signal trans.	566	54	−6.90	0
GO:0023060	signal transmission	2728	173	−6.29	0
GO:0023046	signaling process	2730	173	−6.27	0
GO:0009611	response to wounding	828	68	−6.06	0
GO:0030234	enzyme regulator activity	880	71	−6.02	0
GO:0030695	GTPase regulator activity	437	43	−5.96	0.001
GO:0035466	reg. of sig. pathway	1158	87	−5.96	0.001
GO:0051056	reg. of GTPase mediated sig. trans.	339	36	−5.85	0.001
GO:0023034	intracellular sig. pathway	1708	117	−5.81	0.001
GO:0060589	nucleoside-triphosphatase reg. activity	446	43	−5.73	0.001
GO:0006928	cellular component movement	687	58	−5.61	0.001
GO:0044444	cytoplasmic part	5629	311	−5.60	0.001

TABLE 27-continued

The Top 30 Most Differentially Regulated Gene Ontology Pathways in Macrophage DNA

		Genes		Log^{10}	
GO Category	Category Name	Total	Changed	P-Value	FDR
GO:0007265	Ras protein signal transduction	335	35	−5.54	0.001
GO:0016192	vesicle-mediated transport	743	61	−5.48	0.001
GO:0010876	lipid localization	188	24	−5.44	0.001
GO:0005085	guanyl-nucleotide exchange factor act.	152	21	−5.38	0.001
GO:0035556	intracellular signal transduction	1454	101	−5.31	0.001
GO:0006869	lipid transport	167	22	−5.26	0.001
GO:0010885	reg. of cholesterol storage	12	6	−5.24	0.001
GO:0016477	cell migration	564	49	−5.16	0.001
GO:0042060	wound healing	470	43	−5.15	0.001
GO:0007166	cell surface receptor linked sig. path	1745	116	−5.14	0.001
GO:0005089	Rho guanyl-nucleotide exchange act.	68	13	−5.08	0.001
GO:0051179	localization	3540	207	−5.02	0.002
GO:0001816	cytokine production	240	27	−4.99	0.002

pos. reg. = positive reg.,
sig. = sig.,
trans. = transduction,
FDR = false discovery rate.

Example 9

Summary of AHRR Methylation Patterns in Nicotine Dependence

[0246] In this example we demonstrate the general principle through which substance use status, in this case nicotine use, can be determined by the differential methylation signature. AHRR is a large gene localizing to Chromosome 5 that encodes a competitive antagonist of the aryl hydrocarbon receptor. This large gene (at least 125,000 base pairs) has at least 12 exons, expresses at least 10 different mRNA products and has a large number of CpG enriched regions. In Appendix A, we list the methylation levels at each of the 146 Illumina probes that interrogate this locus for lymphoblasts or pulmonary macrophages (which are derived from peripheral monocytes) derived from either smokers or non-smokers. As Appendix A shows, a number of the probes are differentially methylated in both DNA samples. But in particular, we call attention to the regions (which are in bold typeface) between Probes 58-59 (cg12806681 and cg03991871), Probes 67-78 (cg23576855 and cg05575921) and Probe 84 (cg14817490). As the data demonstrate, the CpG residues in DNA prepared from both smoker lymphoblast and macrophage DNA is significantly hypomethylated at these residues. For any blood derived cell type, it is therefore apparent that by comparing the obtained methylation value for these set of residues for a given patient to that of a non-smoking reference patient, it would be easy to decipher whether the subject is smoking or not. That is to say, if the methylation values at those loci are closer to that obtained from the smokers, then that patient is likely a smoker. Conversely, if the values at those loci are closer to that obtained from the non-smoking reference group, then that patient is likely a non-smoker.

[0247] By this principle, and any of the genes listed in the tables herein or the accompanying Appendices, a skilled practitioner can easily determine with the selected level of surety whether a subject is smoking or non-smoking. By applying the same principle to the differentially methylated genes listed for alcohol or cannabis use, it is also clear that similar determinations can be made for the use of those substances as well.

Example 10

Summary of Illumina Probe Data

[0248] Appendix B is a listing of probes differentially methylated in lymphoblasts of smokers. In this table, a complete listing of all probes differentially methylated at a p-value of $p<0.05$ is provided. Abbreviations: Illumina ID refers to the probe identification according to the Illumina array. "Chrom" and "position" refer to the absolute chromosomal base pair position of the CG dinucleotide according to Build 37 of the human genome. "Methylation" difference indicates whether the methylation at this CpG residue was higher in smokers or lower in smokers.

[0249] Appendix C is a complete listing of all the genes which have at least one Illumina probe differentially methylated in smokers. All gene names are according to the Human Genome Organization (HUGO) convention.

[0250] Appendix D is a listing of probes differentially methylated in lymphoblasts of alcohol drinkers. In this table, a complete listing of all probes differentially methylated at a p-value of $p<0.01$ is provided. Abbreviations: Illumina ID refers to the probe identification according to the Illumina array. "Chrom" and "position" refer to the absolute chromosomal base pair position of the CG dinucleotide according to Build 37 of the human genome. "Methylation" difference indicates whether the methylation at this CpG residue was higher in drinkers or lower in drinkers.

[0251] Appendix E is a complete listing of all the genes which have at least one Illumina probe differentially methylated in drinkers. All gene names are according to the Human Genome Organization (HUGO) convention.

[0252] All publications, patents and patent applications are incorporated herein by reference. While in the foregoing specification this invention has been described in relation to certain embodiments thereof, and many details have been set forth for purposes of illustration, it will be apparent to those skilled in the art that the invention is susceptible to additional embodiments and that certain of the details described herein may be varied considerably without departing from the basic principles of the invention.

[0253] The use of the terms "a" and "an" and "the" and similar referents in the context of describing the invention are to be construed to cover both the singular and the plural, unless otherwise indicated herein or clearly contradicted by context. The terms "comprising," "having," "including," and "containing" are to be construed as open-ended terms (i.e., meaning "including, but not limited to") unless otherwise noted. Recitation of ranges of values herein are merely intended to serve as a shorthand method of referring individually to each separate value falling within the range, unless otherwise indicated herein, and each separate value is incorporated into the specification as if it were individually recited herein. All methods described herein can be performed in any suitable order unless otherwise indicated herein or otherwise clearly contradicted by context. The use of any and all examples, or exemplary language (e.g., "such as") provided herein, is intended merely to better illuminate the invention and does not pose a limitation on the scope of the invention unless otherwise claimed. No language in the specification should be construed as indicating any non-claimed element as essential to the practice of the invention.

[0254] Embodiments of this invention are described herein, including the best mode known to the inventors for carrying out the invention. Variations of those embodiments may become apparent to those of ordinary skill in the art upon reading the foregoing description. The inventors expect skilled artisans to employ such variations as appropriate, and the inventors intend for the invention to be practiced otherwise than as specifically described herein. Accordingly, this invention includes all modifications and equivalents of the subject matter recited in the claims appended hereto as permitted by applicable law. Moreover, any combination of the above-described elements in all possible variations thereof is encompassed by the invention unless otherwise indicated herein or otherwise clearly contradicted by context.

Lengthy table referenced here

US20120108444A1-20120503-T00001

Please refer to the end of the specification for access instructions.

Lengthy table referenced here

US20120108444A1-20120503-T00002

Please refer to the end of the specification for access instructions.

Lengthy table referenced here

US20120108444A1-20120503-T00004

Please refer to the end of the specification for access instructions.

Lengthy table referenced here

US20120108444A1-20120503-T00003

Please refer to the end of the specification for access instructions.

Lengthy table referenced here

US20120108444A1-20120503-T00005

Please refer to the end of the specification for access instructions.

LENGTHY TABLES

The patent application contains a lengthy table section. A copy of the table is available in electronic form from the USPTO web site (http://seqdata.uspto.gov/?pageRequest=docDetail&DocID=US20120108444A1). An electronic copy of the table will also be available from the USPTO upon request and payment of the fee set forth in 37 CFR 1.19(b)(3).

SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 21

<210> SEQ ID NO 1
<211> LENGTH: 30
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic primer

<400> SEQUENCE: 1

ttaaagaatg aaagtattag gttgagagtt 30

<210> SEQ ID NO 2
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic primer

<400> SEQUENCE: 2

atacccactc ttaaaaacca acccc 25

<210> SEQ ID NO 3
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic primer

<400> SEQUENCE: 3

gggtgttgaa ttttgaggag aag 23

<210> SEQ ID NO 4

-continued

<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic primer

<400> SEQUENCE: 4

aaaacacaac tacccaaatc cc 22

<210> SEQ ID NO 5
<211> LENGTH: 28
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic primer

<400> SEQUENCE: 5

ggggagttga tagaagggtt ttttttat 28

<210> SEQ ID NO 6
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic primer

<400> SEQUENCE: 6

tatatctacc tcccccaatc acacc 25

<210> SEQ ID NO 7
<211> LENGTH: 29
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic primer

<400> SEQUENCE: 7

aaagggtggg aaggattttt ttattaatt 29

<210> SEQ ID NO 8
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic primer

<400> SEQUENCE: 8

catcctcaat atccaacttc cccta 25

<210> SEQ ID NO 9
<211> LENGTH: 2050
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic polynucleotide

<400> SEQUENCE: 9

tagtgagggc tggaggctgc gcagacctcg acgggcccta catgacgtca caaaggggcc 60

agaccaagtg gggcagcacc ctgcgaccct gcgatcctgc ctggctcagc cgccttcata 120

tatctgcttc cttaagtcca ctcttgccca gatagctttc agttaaaact aaagaatgaa 180

agcactaggt tgagagccca cgcggctaca cccacgtcta ctcccccact ctcgccaggc 240

-continued

aaccgcgccc cccgcctgca gtggcatcgt ccggccacgc ccagtggcag ggtttccagc 300

gcgagcctgc aggcaggccg ggaaggcgga gccaggccgg cctagagtca cttctccccg 360

cccctgactg ggccgggagc ccggggctgg tctctaagag tgggtaccga gaacagcctg 420

accgtggaga agggctgcgg gaagcagaac accgccccca gcgcccagcg tgctccagaa 480

acatgagcac aaacgcctca gcctccttcc ccggcggcac cggcaccggc accagtaccc 540

gcaccagtac cggcaccggc accagtaccc gcaccagtac cggcaccggc accagtaccc 600

gcaccagtac cggcaccggc accgagcgca aggcggaggg cccgcccgaa gccgggggca 660

caactgccca ggtcccgaac ccggactcca gcttggacga cacctcctac agcctgtccg 720

aatggagcgt ccgttctgag tggcggtccg tctcggatcc gctagccagt tcccagtgga 780

gcacgtcctc aactgccgag gccgcctcct ggagctccag catacactcc ccaatcagca 840

ctaccggtct tagcgagagt actgactccg actccaagag tggcctccgg ggtttcagcg 900

cttacaaccc gagcagtcgg atccccaagt ctaccaccag ctcgaactcc tccgatgggg 960

ccgtcacagc ctccaatcag gacaccggca ttccctgggt attagtaaca ggacctaccc 1020

cgcccgtaaa ctcccccgta gagtcattgc aagggtctgc cttctcctca gggttcagca 1080

ccccacgggg tttggtaaaa ggaccgaccc tgcccccgga ttccaacctg acctcagtgt 1140

ccgactacac ttggatattt gtacggggac ctcctatacc caatgacctt tcgcaagtgt 1200

caatacaagc acctcctaca cccagtaaca cccccgagtg tcagtacaag ggtctgccgc 1260

atcctcagtg tccagcttcc cctggggttt ggtaccagga ccacctctac ccaataacat 1320

ttccccagtg tcgccacaag cacctcctgc accccataac atccccccag tgtcaaggca 1380

ggcgtctacc cccacctcag tgcctgacac tccgcggggt tcaatacaag aacctcctgc 1440

acccagtaat cctttccagc tgccgacaca aggacattct aaacctaata actctcgccg 1500

agtgtcagta caagggtccg ccccgctctc agtgcccagc tccccccggg tatcagctga 1560

aacatcagct ccgcccctgg gcgctcccgg agtatcagca aaagggttcg ccccgcccac 1620

agtgcccggc tccccccggg tatcaaaaga aggatcggct ccgcccccgg gctccccggg 1680

ggagttgata gaagggtcct tcccaccctt tgccgtcccc actcctgtgc ctacgaccca 1740

ggagcgtgtc agccaaagca tggagaatca agagaaggcg agtatcgcgg gccacatgtt 1800

cgacgtagtc gtgatcggag gtggcatttc aggtcagtgt ggaccgtagc ggtggcctgg 1860

gggaccctgg ccagtgaggg gtaggggaac ctacagtagc tcttgtggtg tttgggggtc 1920

tctcatgcat gcgagagtgt agtgtagcca tggcttggcc ccatatcctg cgaggtagga 1980

gtgggggttg tgccagtttt gctggtggtg tgactggggg aggcagacac aataatttta 2040

ctactactac 2050

<210> SEQ ID NO 10
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic primer

<400> SEQUENCE: 10

agtgttggtg tatttatttt aaaa 24

<210> SEQ ID NO 11

-continued

<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic primer

<400> SEQUENCE: 11

tcctaaaaac aaatatcttt caatc 25

<210> SEQ ID NO 12
<211> LENGTH: 27
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic primer

<400> SEQUENCE: 12

taacaatact aatcatttca taaaata 27

<210> SEQ ID NO 13
<211> LENGTH: 27
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic primer

<400> SEQUENCE: 13

agtttagtaa tttggaataa taggttt 27

<210> SEQ ID NO 14
<211> LENGTH: 50
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic probe

<400> SEQUENCE: 14

ttcaggtgcc aggtctggag tgctggtgca cctatctcaa aacgctgtct 50

<210> SEQ ID NO 15
<211> LENGTH: 50
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic probe

<400> SEQUENCE: 15

gcaaacagca gtccagtaac ctggaacaac aggctctgcg aaaccaagga 50

<210> SEQ ID NO 16
<211> LENGTH: 57
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic probe

<400> SEQUENCE: 16

agaaatgaat ggcgttgtca tcgaaaaaac acagactcga ttgtgacaga aataccg 57

<210> SEQ ID NO 17
<211> LENGTH: 50
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:

-continued

<223> OTHER INFORMATION: synthetic probe

<400> SEQUENCE: 17

tgcgcctcca cggaataact gccagccggc acagtgcgag tgagaaaccg 50

<210> SEQ ID NO 18
<211> LENGTH: 50
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic probe

<400> SEQUENCE: 18

ggaaaagaat ccgacgtcgc caacaagcgg tgctaccagg agaaacgcct 50

<210> SEQ ID NO 19
<211> LENGTH: 50
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic probe

<400> SEQUENCE: 19

aaaacacagc tggataaacc gagaaccttc ggagtggttg caccgaaacg 50

<210> SEQ ID NO 20
<211> LENGTH: 50
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic probe

<400> SEQUENCE: 20

gaagcaaccg gcagtgctaa caccgaggag cacctagagc ggcaaaacta 50

<210> SEQ ID NO 21
<211> LENGTH: 1119
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic polynucleotide

<400> SEQUENCE: 21

ggttgaggct gcagtgagct atgcttgtgc cactgcactc cagcctgggg gacacagcca 60

gaccctgtct caaaaaaagt gaaaaaaaaa aaaaagccaa aaaatactgt ggagtggccc 120

tttcttcaag catctgcttg ctttctctaa caccactcat tctattgccc ctactgagct 180

tgaaatgata atgctttctt caggtgccag gtctggagtg ctggtgcacc tatctcaaaa 240

cgctgtctca aaactccaac agggagcacc taacggtact gggtgcaaca ttgctagcac 300

ggagcaaaca gcagtccagt aacctggaac aacaggctct gcgaaaccaa ggactctgac 360

aaagaaaaat tgccaattcc aaacatagcc tgttttagag aaatgaatgg cgttgtcatc 420

gaaaaaacac agactcgatt gtgacagaaa taccgccaca aacgcaggta cagggacagc 480

cgacaccgag aaccaaggga agcggctgag agctgcgcct ccacggaata actgccagcc 540

ggcacagtgc gagtgagaaa ccggccactc catgaaacga ccagtactgc cacggaaaag 600

aatccgacgt cgccaacaag cggtgctacc aggagaaacg cctgcttttg aagaaaacag 660

ccaggaacgc gactgaaaga cacttgctcc caggaagaat tggcatttgt tccaaaacac 720

-continued

```
agctggataa accgagaacc ttcggagtgg ttgcaccgaa acggggtcac ccagcacctc    780

agcgtcctgg gctctagcaa gcctcacaga agcaaccggc agtgctaaca ccgaggagca    840

cctagagcgg caaaactagc agtaatgcca tcgacgaaag gccagttagg ccaaaagaat    900

agaatattta gttccgggaa ttacaggcca gcgcaaacca gacagcataa agctgagggt    960

cagcaaaaca aaaattagga acaatttttt tttaaagggc aagttagctg aaaaacacac   1020

gcacacacaa taaaaacaat acatttggga agattcatca aatgaaaatt caaaactaag   1080

caaaacatgg aaaaatggac tctacaaaga agaaaatgc                          1119
```

What is claimed is:

1. A kit for determining whether a subject is dependent upon a substance, comprising at least one first oligonucleotide probe that is complementary to a sequence that comprises one or more CpG dinucleotides, wherein the at least one first oligonucleotide probe detects either the unmethylated one or more CpG dinucleotides or the methylated one or more CpG dinucleotides, wherein the methylation status of the at least one CpG dinucleotide is associated with the subject's dependence upon the substance.

2. The kit of claim **1**, wherein the sequence that comprises the one or more CpG dinucleotides is contained within the aryl hydrocarbon receptor repressor (AHRR) gene, wherein the substance use disorder is nicotine dependence, and wherein the one or more CpG dinucleotides is at position 373378 of chromosome 5 (CHR5:373378).

3. The kit of claim **1**, wherein the sequence that comprises the one or more CpG dinucleotides is contained within the ARPP19 gene, wherein the substance use disorder is alcohol dependence, and wherein the one or more CpG dinucleotides is at position 50647914 of chromosome 15.

4. The kit of claim **1**, wherein the sequence that comprises the one or more CpG dinucleotides is contained within the AK056486 gene, wherein the substance use disorder is cannabis dependence.

5. The kit of claim **1**, further comprising a solid substrate to which the at least one oligonucleotide probe is attached.

6. The kit of claim **5**, wherein the solid substrate is a polymer, glass, semiconductor, paper, metal, gel or hydrogel.

7. The kit of claim **5**, wherein the solid substrate is a microarray or microfluidics card.

8. The kit of claim **1**, further comprising at least one second oligonucleotide probe that is complementary to the sequence that comprises one or more CpG dinucleotides, wherein the at least one second oligonucleotide probe detects either the unmethylated one or more CpG dinucleotides or the methylated one or more CpG dinucleotides that is not detected by the at least one first oligonucleotide probe.

9. The kit of claim **1**, further comprising a control oligonucleotide probe.

10. A method for determining whether a subject is dependent upon a substance comprising:

(a) providing a biological sample from the subject;

(b) contacting DNA from the biological sample with bisulfite under alkaline conditions;

(c) contacting the bisulfite-treated DNA with at least one first oligonucleotide probe that is complementary to a sequence that comprises one or more CpG dinucleotides, wherein the at least one first oligonucleotide probe detects either the unmethylated one or more CpG dinucleotides or the methylated one or more CpG dinucleotides,

wherein the methylation status of the at least one CpG dinucleotide is associated with the subject's dependence upon the substance.

11. The method of claim **10**, wherein the biological sample is peripheral blood.

12. The method of claim **10**, wherein the biological sample is predominantly lymphocytes.

13. The method of claim **10**, further comprising contacting the bisulfite-treated DNA with at least one second oligonucleotide probe that is complementary to the region that comprises one or more CpG dinucleotides, wherein the at least one second oligonucleotide probe detects either the unmethylated one or more CpG dinucleotides or the methylated one or more CpG dinucleotides that is not detected by the at least one first oligonucleotide probe.

14. The method of claim **13**, further comprising determining the ratio of methylated one or more CpG dinucleotides to unmethylated one or more CpG dinucleotides.

15. The method of claim **10**, further comprising an amplifying step after the contacting step.

16. The method of claim **10**, further comprising a sequencing step after the contacting step.

17. The method of claim **10**, wherein the substance is nicotine, alcohol, or cannabis.

18. A method for determining whether a subject is dependent upon a substance comprising:

(a) providing a biological sample from the subject;

(b) contacting DNA from the biological sample with an antibody that binds to methylated cytosine nucleotides;

(c) comparing the DNA sequences bound by the antibody to corresponding unmethylated reference DNA sequences,

wherein the methylation status of the DNA sequences from the biological sample relative to the reference DNA sequences is associated with the subject's dependence upon the substance.

19. The method of claim **18**, wherein at least 20 different DNA sequences bound by the antibody are compared to the reference DNA sequences.

* * * * *

专利名称(译)	用于检测物质使用障碍的易感性的组合物和方法		
公开(公告)号	US20120108444A1	公开(公告)日	2012-05-03
申请号	US13/284425	申请日	2011-10-28
[标]申请(专利权)人(译)	PHILIBERT ROBERT 马丹ANUP		
申请(专利权)人(译)	PHILIBERT ROBERT 马丹ANUP		
当前申请(专利权)人(译)	PHILIBERT ROBERT 马丹ANUP		
[标]发明人	PHILIBERT ROBERT MADAN ANUP		
发明人	PHILIBERT, ROBERT MADAN, ANUP		
IPC分类号	C40B20/08 C08G83/00 C12Q1/68 G01N33/53 C07H21/04 C40B40/06		
CPC分类号	C12Q1/6869 C12Q1/6883 C12Q2600/112 C12Q2600/154 C12Q2600/124 Y10T436/143333 C12Q2523/125		
优先权	61/173274 2009-04-28 US		
其他公开文献	US8637652		
外部链接	Espacenet USPTO		

摘要(译)

本发明提供筛选试剂盒，组合物和诊断方法，用于通过测定来自受试者的生物样品的核酸甲基化谱来确定受试者是否具有物质使用障碍的倾向或可能性，其中给定的概况表明受试者具有物质使用障碍的倾向。