

(51) International Patent Classification:
C12Q 1/68 (2006.01)

(21) International Application Number:

PCT/US2012/038552

(22) International Filing Date:

18 May 2012 (18.05.2012)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

61/488,277

20 May 2011 (20.05.2011)

US

(71) Applicant (for all designated States except US): **THE REGENTS OF THE UNIVERSITY OF CALIFORNIA** [US/US]; 1111 Franklin Street, 12th Floor, Oakland, CA 94607 (US).

(72) Inventors; and

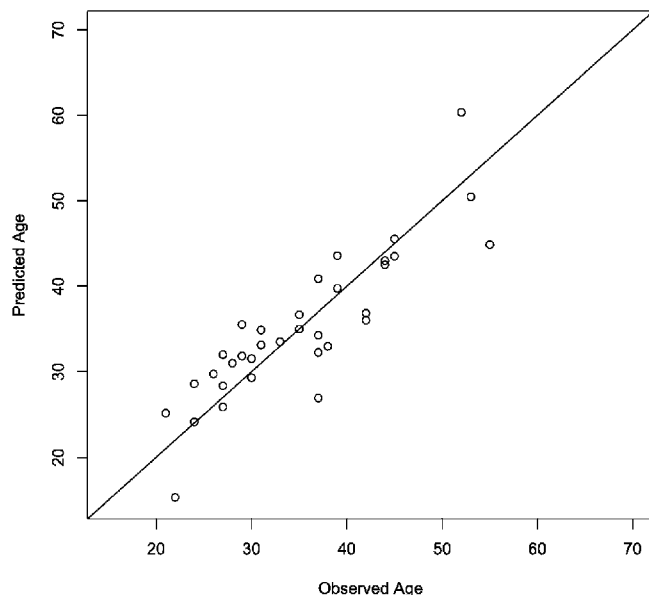
(75) Inventors/Applicants (for US only): **VILAIN, Eric** [US/US]; 1075 Hanley Avenue, Los Angeles, CA 90049 (US). **BOCKLANDT, Sven** [BE/US]; 1151 Laveta Ter, Los Angeles, CA 90026 (US). **HORVATH, Steve**[US/US]; 149 S. Barrington Avenue, #713, Los Angeles, CA 90049 (US). **SINSHEIMER, Janet** [US/US]; 10827 Cushman Avenue, Los Angeles, CA 90064 (US).(74) Agent: **WOOD, William, J.**; Gates & Cooper LLP, 6701 Center Drive West, Suite 1050, Los Angeles, CA 90045 (US).

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

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK,

[Continued on next page]

(54) Title: METHOD TO ESTIMATE AGE OF INDIVIDUAL BASED ON EPIGENETIC MARKERS IN BIOLOGICAL SAMPLE

**FIG. 11**

(57) Abstract: The invention provides methods and materials that use observation of DNA characteristics to obtain information relating to the age of individuals. The instant disclosure identifies 88 sites in or near 80 genes for which the degree of cytosine methylation in epithelial and/or white blood cells is significantly correlated with age. In illustrative embodiments of the invention, cytosine methylation patterns the promoters of the EDARADD, TOMIL1, and NPTX2 genes are used to predict the age of an individual with a high degree of accuracy.



EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU,
LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK,
SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ,
GW, ML, MR, NE, SN, TD, TG).

- *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*
- *with sequence listing part of description (Rule 5.2(a))*

Published:

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

METHOD TO ESTIMATE AGE OF INDIVIDUAL BASED ON EPIGENETIC MARKERS IN BIOLOGICAL SAMPLE

REFERENCE TO RELATED APPLICATIONS

5 This application claims priority under Section 119(e) from U.S. Provisional Application Serial No. 61/488,277, filed May 20, 2011, the contents of which are incorporated herein by reference.

BACKGROUND OF THE INVENTION

10 1. Field of the Invention.

 This invention relates to methods of detecting and analyzing patterns of cytosine methylation in genomic DNA.

2. Description of Related Art.

15 Throughout development, cells and tissues differentiate and change as an organism ages. These changes include alterations to telomeres, accumulation of DNA mutations, decay of cellular and organ structures, and changes in gene expression (see, e.g. Goyns MH (2002) *Mech Ageing Dev* 123: 791-799). Both differentiation of tissues and aging effects are at least partially caused by chemical modifications of the genome, such as DNA methylation. In particular, the genomic DNA of higher eukaryotes contains modified nucleosides including 5-methyl cytosines. This modification is usually found as part of the dinucleotide CpG.

 DNA methylation is an epigenetic determinant of gene expression. Patterns of CpG methylation are heritable, tissue specific, and correlate with gene expression.

25 The consequence of methylation is usually gene silencing. DNA methylation also correlates with other cellular processes including embryonic development, chromatin structure, genomic imprinting, somatic X-chromosome inactivation in females, inhibition of transcription and transposition of foreign DNA and timing of DNA replication. When a gene is highly methylated it is less likely to be expressed. Thus

30 the identification of sites in the genome containing 5-mC is important in

understanding cell-type specific programs of gene expression and how gene expression profiles are altered during both normal development and diseases such as cancer. Mapping of DNA methylation patterns is important for understanding diverse biological processes such as the regulation of imprinted genes, X chromosome inactivation, and tumor suppressor gene silencing in human cancers.

Several studies have investigated the epigenetic state of a small number of selected genes or CpG islands in subjects of varying age or have measured the global changes in DNA methylation with increasing age (see, e.g. Boks MP, et al. (2009) *PLoS One* 4: e6767; and Fraga MF, et al. (2005) *Proc Natl Acad Sci USA* 102: 10604-10609). Recently, unbiased genome-wide studies have documented age effects on DNA methylation in cultured cells, mice, and humans (see, e.g. Bork S, et al. (2009) *Aging Cell* 9: 54-63; Maegawa S, et al. (2010) *Genome Res* 20: 332-340; and Teschendorff AE, et al. (2010) *Genome Res* 20: 440-446; Gronniger E, et al. (2010) *PLoS Genet* 6: e1000971; Rakyan VK, et al. (2010) *Genome Res* 20: 434-439). In these reports, the subject's were of a limited age range, and the continuity of the age related changes was not defined. Consequently, estimating the age of an individual by observing methylation patterns in their genomic DNA has not been possible.

Methods for estimating the age of an individual by observing methylation patterns in genomic DNA obtained from a biological sample have a number of applications. For example, the characterization of biological materials is one of the most important methods for identification of individuals in forensic medicine and/or in criminal investigations (see, e.g. van Oorschot et al., *Investigative Genetics* 2010, 1:14; and Thompson et al., *Methods Mol Biol.* 2012;830:3-16). When analyzing biological materials found at a crime scene, common procedures include DNA analysis techniques such as DNA fingerprinting to specifically identify the individual from which the biological material was derived. New DNA analysis techniques, for example those that can be used to predict an approximate age of an individual, are desirable.

SUMMARY OF THE INVENTION

As noted above, the genetic material in individuals is observed to change as they age. The disclosure provided herein includes experiments examining such changes by measuring patterns of DNA methylation in a population of individuals ranging in age from 18 to 70. These experiments identified 88 specific regions within human genomic DNA where the amount of DNA methylation correlates with the age of the individual. The invention disclosed herein includes methodologies where an amount of DNA methylation can be correlated with a predicted age. In certain working embodiments of the invention, the average absolute difference between a predicted age and actual age was 3.5 years. Some embodiments of the invention can be used, for example, by forensic scientists to estimate the age of an individual from a biological sample found at a crime scene. Other embodiments of the invention can also be used by medical personnel to assess, for example, an individual's risk for one or more age related pathologies.

The invention disclosed herein has a number of embodiments. One embodiment of the invention is a method of obtaining information useful to predict an age of an individual by observing the methylation status of one or more specific loci in the genomic DNA. Such embodiments comprise the steps of obtaining genomic DNA from white blood cells or epithelial cells derived from the individual; and then observing cytosine methylation of one or more (and typically two, three, four or more) CG loci in the genomic DNA selected from the group consisting of CG locus designation: cg27553955, cg12799895, cg21296230, cg04084157, cg25148589, cg17861230, cg03734874, cg18236477, cg27320127, cg07621046, cg14456683, cg21801378, cg00107187, cg05508084, cg07533148, cg03975694, cg19945840, cg00399483, cg06291867, cg24826867, cg06092815, cg19885761, cg24199834, cg04528819, cg13434842, cg02008154, cg25044651, cg02994956, cg20366906, cg27389185, cg14826456, cg12111714, cg12457773, cg11981599, cg00059225, cg06572160, cg08668790, cg12782180, cg20134215, cg13921352, cg13614181, cg23563234, cg00201234, cg21992250, cg00911351, cg19594666, cg15425280,

cg06156376, cg19831077, cg27409364, cg15747595, cg02154186, cg20616414, cg10235817, cg16232126, cg24646414, cg17241310, cg10031651, cg23290344, cg06908778, cg02844545, cg14614211, cg25511429, cg15201635, cg19246110, cg06760035, cg13603171, cg13282837, cg20792062, cg02228185, cg13547237, 5 cg09809672, cg01293143, cg05822532, cg07408456, cg08468689, cg27210390, cg01820374, cg17589341, cg19761273, cg03440846, cg08909157, cg11136562, cg16464322, cg08872742, cg18328933, cg15784615 and cg23282949. In this way, information useful to predict the age of the individual is obtained.

Related embodiments of the invention include methods of obtaining 10 information useful to predict an age of an individual by observing methylation patterns one or more specific gene sequences. Illustrative embodiments of this aspect of the invention comprise the steps of obtaining a biological sample derived from an individual comprising genomic DNA from white blood cells or epithelial cells; and then observing a pattern of cytosine methylation occurring on at least one genomic 15 DNA sequence selected from the group consisting of Edaradd (SEQ ID NO: 1), TomL1 (SEQ ID NO: 2), NPTX2 (SEQ ID NO: 3) and ELN (SEQ ID NO: 4). In this way, information useful to predict the age of the individual is obtained. Certain embodiments of the invention observe methylation patterns in multiple gene sequences. Optionally, for example, cytosine methylation is observed in both 20 Edaradd (SEQ ID NO: 1) and TomL1 (SEQ ID NO: 2); and/or both Edaradd (SEQ ID NO: 1) and NPTX2 (SEQ ID NO: 3). In some embodiments of the invention, cytosine methylation is observed in at least two genomic DNA sequences including ELN (SEQ ID NO: 4).

Embodiments of the invention comprise correlating patterns of cytosine 25 methylation with a predicted chronological age of the individual. For example, as disclosed herein, the methylation of certain sites such as the promoters of the EDARADD, TOM1L1, and NPTX2 genes increases with age in a linear manner over a range of five decades. In this context, certain working embodiments use well known mathematical modeling techniques (e.g. regression analyses) to correlate an

observed pattern of cytosine methylation a predicted age of the individual. In certain embodiments the methods of the invention can be used to provide valuable information in forensic investigations (e.g. where the identity of the individual from which the DNA is derived is unknown). Such embodiments of the invention can be combined with other forensic analysis procedures, for example by also performing a DNA fingerprinting analysis on the genomic DNA. In other embodiments of the invention, one can compare the predicted age of the individual with the actual chronological age of the individual, for example as part of a diagnostic procedure for an age associated pathology (e.g. one that compares an individual's chronological age with an apparent biological age in view of their DNA methylation patterns). Such methods can be useful in clinical interventions that are predicated on an epigenetic bio-age rather than an actual chronological age.

As discussed in detail below, the methods of the invention can be adapted for use with a variety of art accepted processes. For example, in certain embodiments of the invention, a bisulfite conversion process is performed so that cytosine residues in the genomic DNA are transformed to uracil, while 5-methylcytosine residues in the genomic DNA are not transformed to uracil. Optionally, the genomic DNA is transformed from its natural state via amplification by a polymerase chain reaction process. In certain embodiments of the invention, the genomic DNA is hybridized to a complimentary sequence (e.g. a synthetic polynucleotide sequence) that is coupled to a matrix (e.g. one disposed within a microarray).

Embodiments of the invention also provide articles of manufacture and kits for obtaining information useful to determine the age of an individual. In an illustrative embodiment, the kit includes a plurality of primers or probes specific for at least two genomic DNA sequences in a biological sample, wherein the genomic DNA sequences comprise one or more of the CG loci in the genomic DNA identified in the Table shown in Figure 4. Such kits of the invention can further include additional reagents, for example a reagent used in a genomic DNA polymerization process, a reagent used in a genomic DNA hybridization process, and/or a reagent used in a

genomic DNA bisulfite conversion process. Optionally, the kit comprises a plurality of primer sets for amplifying at least two genomic DNA sequences. In certain embodiments of the invention, the kit comprises primer sets for amplifying at least two genomic DNA sequences in the biological sample selected from the group consisting of Edaradd (SEQ ID NO: 1), TomL1 (SEQ ID NO: 2), NPTX2 (SEQ ID NO: 3) and ELN (SEQ ID NO: 4). In some embodiments of the invention, the kit further comprises a probe or primer used to perform a DNA fingerprinting analysis.

Other objects, features, and advantages of the present invention will become apparent to those skilled in the art from the following detailed description. It is to be understood, however, that the detailed description and specific examples, while indicating some embodiments of the present invention are given by way of illustration and not limitation. Many changes and modifications within the scope of the present invention may be made without departing from the spirit thereof, and the invention includes all such modifications.

15

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 illustrates the detection of gene co-methylation modules in human saliva in twins. (a) Branches of the hierarchical cluster tree define five co-methylation modules which are assigned a color as can be seen from the first color band underneath the tree. Probes that could not be clustered into one of these modules were coded. Every probe represents a line in the hierarchical cluster tree. Distance between two probes is shown as height on the y-axis. The second band encodes the age relationships of each gene. Genes with positive age correlations are highlighted. (b) Barplots showing age relationships of modules. Specifically, the y-axis shows the mean Student T-statistic testing whether the methylation status of a probe is correlated with age. Note that the module is enriched for probes that have a significant positive correlation with age. A t-statistic value of 2 or higher indicates a significant correlation ($p < 0.05$).

25

Figure 2 illustrates percentage methylation versus age for three markers validated in three sample sets. Original twin samples are blue, male control samples are red, female control samples green. Linear trendlines are shown in the colors of the individual sample sets a) Edaradd $r=-0.81$ (twins), $r=-0.73$ (male controls), $r=-0.75$ (female controls) b) NPTX2 $r=0.52$ (twins), $r=0.79$ (male controls), $r=0.03$ (female controls) c) Tom1L1 $r=-0.70$ (twins), $r=-0.49$ (male controls), $r=-0.24$ (female controls).

Figure 3 illustrates predicted versus observed age of all subjects using a leave-one-out model. A multivariate regression model was fit on all but one sample and its predicted age (y-axis) was related to the truly observed age of the left out sample (x-axis). The predicted values are highly correlated with the observed ages ($r=0.83$, $p=2.2 \times 10^{-16}$, $n=66$), and the average absolute difference between the predicted and the observed age is 5.2 years.

Figures 4A-4E illustrate 88 loci that are determined to be significantly correlated with age. TargetID represents the exact Illumina probe on the array, Chr: chromosome number, Gene_ID: NCBI Gene database locator, Symbol: gene name, r: correlation coefficient, p-value: significance of correlation, q-value: significance corrected for multiple comparisons.

Figure 5 illustrates array probes found to be positively correlated with age in blood (see e.g. Rakyan VK, et al. Human aging-associated DNA hypermethylation occurs preferentially at bivalent chromatin domains. *Genome Res* 20: 434-439) and in saliva (present study).

Figure 6 illustrates disease and molecular function categories significantly enriched in ingenuity analysis.

Figure 7 illustrates unsupervised hierarchical clustering of all samples. The y-axis shows distance between samples. Each twin pair is color coded. Row "Pair" shows that the majority of twin pairs cluster together. Samples were divided in the oldest and youngest half and coded dark and light blue. Row "Age" shows that samples of similar age group did not cluster together. The different arrays were each

color coded as well, and row “Array” shows that samples hybridized together do not cluster together, suggesting that variations in hybridization do contribute to the data analysis.

Figure 8 illustrates predicted versus observed age of all male subjects using a leave-one-out model. A multivariate regression model was fit on all but one sample and its predicted age (y-axis) was related to the truly observed age of the left out sample (x-axis). The predicted values are highly correlated with the observed outcomes ($r=0.83$, $p=3.3 \times 10^{-13}$, $n=47$), and the average absolute difference between the predicted and the observed age is 5.3 years.

Figure 9 illustrates predicted versus observed age of all female subjects using a leave-one-out model. A multivariate regression model was fit on all but one sample and its predicted age (y-axis) was related to the truly observed age of the left out sample (x-axis). The predicted values are highly correlated with the observed outcomes ($r=0.75$, $p=2.4 \times 10^{-4}$, $n=19$), and the average absolute difference between the predicted and the observed age is 6.2 years.

Figure 10 illustrates predicted versus observed age of all non-twin subjects using a leave-one-out model. A multivariate regression model was fit on all but one sample and its predicted age (y-axis) was related to the truly observed age of the left out sample (x-axis). The predicted values are highly correlated with the observed outcomes ($r=0.85$, $p=1.701 \times 10^{-13}$, $n=45$) and the average absolute difference between the predicted and the observed age is 5.3 years.

Figure 11 illustrates predicted versus observed age of all twin subjects using a leave-one-out model. A multivariate regression model was fit on data of previously used markers plus the methylation value at the ELN gene, on microarray data, for all but one sample and its predicted age (y-axis) was related to the truly observed age of the left out sample (x-axis). The predicted values are highly correlated with the observed outcomes ($r=0.87$, $p=2.2 \times 10^{-11}$, $n=34$), and the average absolute difference between the predicted and the observed age is 3.5 years.

DETAILED DESCRIPTION OF THE INVENTION

Many of the techniques and procedures described or referenced herein are well understood and commonly employed by those skilled in the art. All publications mentioned herein (e.g. U.S. Patent No. 7,700,324, U.S. Patent Application Nos. 5 20040132026 and 20060292585, Berdasco et al., Aging Cell (2012) 11, pp181–186; Bockland et al., et al. (2011) PLoS ONE 6(6): e14821. doi:10.1371/journal.pone.0014821; Koch et al., (2011) AGING, Vol 3, No 10 , pp 1,018-1,027) are incorporated herein by reference to disclose and describe aspects, methods and/or materials in connection with the cited publications. Publications cited 10 herein are cited for their disclosure prior to the filing date of the present application. Nothing here is to be construed as an admission that the inventors are not entitled to antedate the publications by virtue of an earlier priority date or prior date of invention. Further the actual publication dates may be different from those shown and require independent verification. In the description of embodiments, reference may be made 15 to the accompanying figures which form a part hereof, and in which is shown by way of illustration a specific embodiment in which the invention may be practiced. It is to be understood that other embodiments may be utilized and structural changes may be made without departing from the scope of the present invention.

Unless otherwise defined, all terms of art, notations and other scientific terms 20 or terminology used herein are intended to have the meanings commonly understood by those of skill in the art to which this invention pertains. In some cases, terms with commonly understood meanings are defined herein for clarity and/or for ready reference, and the inclusion of such definitions herein should not necessarily be construed to represent a substantial difference over what is generally understood in 25 the art.

The term "genome" or "genomic" as used herein is all the genetic material in the chromosomes of an organism. DNA derived from the genetic material in the chromosomes of a particular organism is genomic DNA.

The term "epigenetic" as used herein means relating to, being, or involving a modification in gene expression that is independent of DNA sequence. Epigenetic factors include modifications in gene expression that are controlled by changes in DNA methylation and chromatin structure. For example, methylation patterns are known to correlate with gene expression.

The term "nucleic acids" as used herein may include any polymer or oligomer of pyrimidine and purine bases, preferably cytosine, thymine, and uracil, and adenine and guanine, respectively. See Albert L. Lehninger, PRINCIPLES OF BIOCHEMISTRY, at 793-800 (Worth Pub. 1982). The present invention contemplates any deoxyribonucleotide, ribonucleotide or peptide nucleic acid component, and any chemical variants thereof, such as methylated, hydroxymethylated or glucosylated forms of these bases, and the like. The polymers or oligomers may be heterogeneous or homogeneous in composition, and may be isolated from naturally-occurring sources or may be artificially or synthetically produced. In addition, the nucleic acids may be DNA or RNA, or a mixture thereof, and may exist permanently or transitionally in single-stranded or double-stranded form, including homoduplex, heteroduplex, and hybrid states.

The terms "oligonucleotide" and "polynucleotide" as used herein refers to a nucleic acid ranging from at least 2, preferable at least 8, and more preferably at least 20 nucleotides in length or a compound that specifically hybridizes to a polynucleotide. Polynucleotides of the present invention include sequences of deoxyribonucleic acid (DNA) or ribonucleic acid (RNA) which may be isolated from natural sources, recombinantly produced or artificially synthesized and mimetics thereof.

The term "probes" as used herein are oligonucleotides capable of binding in a base-specific manner to a complementary strand of nucleic acid. The term "probe" as used herein refers to a surface-immobilized molecule that can be recognized by a particular target as well as molecules that are not immobilized and are coupled to a detectable label. The term "label" as used herein refers, for example, to colorimetric

(e.g. luminescent) labels, light scattering labels or radioactive labels. Fluorescent labels include, inter alia, the commercially available fluorescein phosphoramidites such as Fluoreprime (Pharmacia), Fluoredate (Millipore) and FAM (ABI). See U.S. Pat. No. 6,287,778.. See U.S. Pat. No. 6,582,908 for an example of arrays having all
5 possible combinations of probes with 10, 12, and more bases.

The term "primer" as used herein refers to a single-stranded oligonucleotide capable of acting as a point of initiation for template-directed DNA synthesis under suitable conditions for example, buffer and temperature, in the presence of four different nucleoside triphosphates and an agent for polymerization, such as, for
10 example, DNA or RNA polymerase or reverse transcriptase. The length of the primer, in any given case, depends on, for example, the intended use of the primer, and generally ranges from 15 to 30 nucleotides. A primer need not reflect the exact sequence of the template but must be sufficiently complementary to hybridize with such template. The primer site is the area of the template to which a primer
15 hybridizes. The primer pair is a set of primers including a 5' upstream primer that hybridizes with the 5' end of the sequence to be amplified and a 3' downstream primer that hybridizes with the complement of the 3' end of the sequence to be amplified.

The term "complementary" as used herein refers to the hybridization or base pairing between nucleotides or nucleic acids, such as, for instance, between the two
20 strands of a double stranded DNA molecule or between an oligonucleotide primer and a primer binding site on a single stranded nucleic acid to be sequenced or amplified. Complementary nucleotides are, generally, A and T (or A and U), or C and G. Two single stranded RNA or DNA molecules are said to be complementary when the nucleotides of one strand, optimally aligned and compared and with appropriate
25 nucleotide insertions or deletions, pair with at least about 80% of the nucleotides of the other strand, usually at least about 90% to 95%, and more preferably from about 98 to 100%. Alternatively, complementarity exists when an RNA or DNA strand will hybridize under selective hybridization conditions to its complement. Typically, selective hybridization will occur when there is at least about 65% complementary

over a stretch of at least 14 to 25 nucleotides, preferably at least about 75%, more preferably at least about 90% complementary. See, M. Kanehisa, Nucleic Acids Res. 12:203 (1984), incorporated herein by reference.

5 The term "hybridization" as used herein refers to the process in which two single-stranded polynucleotides bind non-covalently to form a stable double-stranded polynucleotide; triple-stranded hybridization is also theoretically possible. Factors that can affect the stringency of hybridization, including base composition and length of the complementary strands, presence of organic solvents and extent of base mismatching, the combination of parameters is more important than the absolute
10 measure of any one alone. Hybridization conditions suitable for microarrays are described in the Gene Expression Technical Manual, 2004 and the GeneChip Mapping Assay Manual, 2004, available at Affymetrix.com.

The term "array" or "microarray" as used herein refers to an intentionally created collection of molecules which can be prepared either synthetically or
15 biosynthetically (e.g. Illumina HumanMethylation27 microarrays). The molecules in the array can be identical or different from each other. The array can assume a variety of formats, for example, libraries of soluble molecules; libraries of compounds tethered to resin beads, silica chips, or other solid supports.

The term "solid support", "support", and "substrate" as used herein are used
20 interchangeably and refer to a material or group of materials having a rigid or semi-rigid surface or surfaces. In many embodiments, at least one surface of the solid support will be substantially flat, although in some embodiments it may be desirable to physically separate synthesis regions for different compounds with, for example, wells, raised regions, pins, etched trenches, or the like. According to other
25 embodiments, the solid support(s) will take the form of beads, resins, gels, microspheres, or other geometric configurations. See U.S. Pat. No. 5,744,305 for exemplary substrates.

DESCRIPTION OF ILLUSTRATIVE ASPECTS OF THE INVENTION

From the moment of conception, we begin to age. A decay of cellular structures, gene regulation, and DNA sequence ages cells and organisms. For example, DNA methylation patterns change with increasing age and contribute to age related disease. Methylation in promoter regions is generally accompanied by gene silencing and loss of methylation or loss of the proteins that bind to the methylated CpG can lead to diseases in humans, for example, Immunodeficiency Craniofacial Syndrome and Rett Syndrome, Bestor (2000) Hum. Mol. Genet. 9:2395-2402. DNA methylation may be gene-specific and occurs genome-wide.

A variety of methods for detecting methylation status or patterns have been described in, for example U.S. Pat. Nos. 6,214,556, 5,786,146, 6,017,704, 6,265,171, 6,200,756, 6,251,594, 5,912,147, 6,331,393, 6,605,432, and 6,300,071 and US Patent Application publication Nos. 20030148327, 20030148326, 20030143606, 20030082609 and 20050009059, each of which are incorporated herein by reference. Other array based methods of methylation analysis are disclosed in U.S. patent application Ser. No. 11/058,566. For a review of some methylation detection methods, see, Oakeley, E. J., Pharmacology & Therapeutics 84:389-400 (1999). Available methods include, but are not limited to: reverse-phase HPLC, thin-layer chromatography, SssI methyltransferases with incorporation of labeled methyl groups, the chloroacetaldehyde reaction, differentially sensitive restriction enzymes, hydrazine or permanganate treatment (m5C is cleaved by permanganate treatment but not by hydrazine treatment), sodium bisulfite, combined bisulphate-restriction analysis, and methylation sensitive single nucleotide primer extension.

As disclosed herein, a number of locations have been identified in the human genome for which the percentage of DNA methylation is linearly correlated with age. By measuring this DNA methylation at the disclosed locations, for example, at just a few of the 3 billion letters in a person's genome, the present invention is able to make accurate estimates of a person's chronological age. While previous studies have shown that DNA methylation in certain parts of the genome changes with age, the

experiments disclosed herein are the first to identify loci where methylation is continuously correlated with age, over a range of at least 5 decades. The strength of this finding allows the present invention to accurately predict an individual's age. In certain embodiments of the invention, the link between age and this chemical change in the DNA is so strong that it is possible to estimate how old someone is by examining, for example, just two spots in the genome of the individual. Aspects of the invention have been published in a peer reviewed technical journal (see Bockland et al., et al. (2011) PLoS ONE 6(6): e14821. doi:10.1371/journal.pone.0014821). In addition, certain aspects of this research have been confirmed by other studies (see, e.g. Koch et al., (2011) AGING, Vol 3, No 10 , pp 1,018-1,027)

As described in detail in the Example below, specific age related methylation markers have been identified and validated using further assays and additional samples. As discussed below, illustrative age prediction analyses models were also designed and tested, for example using a leave-one-out analysis, where one subject from the model is systematically removed and the model is used to predict the subject's age. As the real age of this subject is already known, such methods provided ways to validate various model designs. Such age predictive models can be applied in a variety of contexts. For instance, the ability to predict an individual's age (e.g. to an average accuracy of 3 to 5 years) can be used by forensic scientists to estimate a suspect's age based on a biological sample alone. In embodiments of the invention designed for forensic use, a practitioner could, for example, submit a biological sample to a lab. In the lab, DNA prepared from the sample could then be analyzed so that the percentage of methylation at one or more of the loci identified herein is determined. These results could be input in a regression model such as those disclosed herein, in order to predict the age of the suspect.

Age prediction methodologies are also relevant to healthcare applications. For example, significant DNA methylation differences are known to be associated with specific age-related disorders, for example in comparisons between the brains of people diagnosed with late-onset Alzheimer's disease and brains from controls. In

this context, the identification of specific loci highly correlated with age can be used to enhance the understanding of aging in health and disease. In certain embodiments of the invention, age prediction methodologies can be used as part of clinical interventions tailored for patients based on their “bio-age”—a result of the interaction of genes, environment, and time—rather than their chronological age. For example, if a person's predicted age is higher than their real age, specific interventions could be designed to return the genome to a “younger” state. Age prediction methodologies can also pave the way for interventions based on specific epigenetic marks associated with disease, as occurs in certain cancer treatments.

The invention disclosed herein has a number of embodiments. As shown by the experiments disclosed herein, 88 sites have been identified in or near 80 genes for which the degree of cytosine methylation significantly correlates with age, for example, in biological samples obtained from the saliva of 34 male identical twin pairs between 21 and 55 years of age. There are more than 28 million CpG loci in the human genome. Consequently, certain loci are given unique identifiers such as those found in the Illumina CpG loci database (see, e.g. Technical Note: Epigenetics, *CpG Loci Identification* ILLUMINA Inc. 2010). These CG locus designation identifiers are used herein. In this context, one embodiment of the invention is a method of obtaining information useful to determine an age of an individual by observing the methylation status of one or more specific GC loci that are identified in Figure 4. A related embodiment of the invention is a method of obtaining information useful to determine an age of an individual by observing the methylation status of one or more cytosines in genomic DNA that is within 1 kilobase on either side of the specific GC loci that are identified in Figure 4 (see, e.g., Table 1 below).

Embodiments of the invention comprise the steps of obtaining genomic DNA from leukocytes or epithelial cells derived from the individual and then observing cytosine methylation of one or more (and typically two, three, four or more) CG loci in the genomic DNA. Leukocytes and/or epithelial cells can be derived from a variety of biological samples (e.g. saliva, blood, semen, skin and the like). Such samples can,

for example, be treated with a reagent effective for opening the cells, fluids, tissues, cell membranes of the sample, and for exposing the nucleic acid(s). Methods for purifying or partially purifying nucleic acid from a sample are well known in the art (e.g., Sambrook et al., *Molecular Cloning: a Laboratory Manual*. Cold Spring Harbor Press, 1989, herein incorporated by reference).

In illustrative embodiments of the invention, methylation is observed in one or more CG loci selected from the group consisting of CG locus designation: cg27553955, cg12799895, cg21296230, cg04084157, cg25148589, cg17861230, cg03734874, cg18236477, cg27320127, cg07621046, cg14456683, cg21801378, cg00107187, cg05508084, cg07533148, cg03975694, cg19945840, cg00399483, cg06291867, cg24826867, cg06092815, cg19885761, cg24199834, cg04528819, cg13434842, cg02008154, cg25044651, cg02994956, cg20366906, cg27389185, cg14826456, cg12111714, cg12457773, cg11981599, cg00059225, cg06572160, cg08668790, cg12782180, cg20134215, cg13921352, cg13614181, cg23563234, cg00201234, cg21992250, cg00911351, cg19594666, cg15425280, cg06156376, cg19831077, cg27409364, cg15747595, cg02154186, cg20616414, cg10235817, cg16232126, cg24646414, cg17241310, cg10031651, cg23290344, cg06908778, cg02844545, cg14614211, cg25511429, cg15201635, cg19246110, cg06760035, cg13603171, cg13282837, cg20792062, cg02228185, cg13547237, cg09809672, cg01293143, cg05822532, cg07408456, cg08468689, cg27210390, cg01820374, cg17589341, cg19761273, cg03440846, cg08909157, cg11136562, cg16464322, cg08872742, cg18328933, cg15784615 and cg23282949. In this way, information useful to determine the age of the individual is obtained.

As a confirmation of the significance of the CpG loci data obtained using Illumina HumanMethylation27 microarrays, certain sites were validated, for example those in the promoters of the EDARADD, TOM1L1, and NPTX2 genes. The results were then replicated in a general population sample of 31 males and 29 females between 18 and 70 years of age. These validation studies confirm, for example, that the methylation of sites including those in the promoters of the EDARADD,

TOM1L1, and NPTX2 genes, is linear with age over a range of at least five decades. Table 1 below shows the surrounding sequences and the exact location of certain CpGs of interest in the Edaradd, Tom1L1, NPXT2 and ELN sequences. Different font emphasis is used to show the location of the PCR primers, of the sequencing
5 primers (for two genes, the third gene was measured using different methods as discussed in the text) as well as the CpGs found to be correlated with age. Font emphasis is also used for other CpGs that can be read in the same experiment and be correlated with age. As shown in Table 1, NPTX2 has more than one CpG that are underlined (CpG 1, 2, 4 and 6 starting from the sequencing primer). These have been
10 shown to be correlated with age as well. The methylation status of one of these is highly predictive of the methylation of the others nearby, so any one of them can be used in the predictive models disclosed herein. Those of skill in this art understand that the CpGs in the vicinity of the relevant ones can further be examined in embodiments of the invention (e.g. because methylation status is sometimes clustered
15 on those CpGs for hundreds or thousands of basepairs away).

Embodiments of the invention include methods of obtaining information useful to determine an age of an individual by observing methylation patterns one or more specific gene sequences. Such embodiments comprise the steps of obtaining a biological sample derived from an individual comprising genomic DNA from white
20 blood cells or epithelial cells; and then observing a pattern of cytosine methylation of at least one genomic DNA sequence selected from the group consisting of Edaradd (SEQ ID NO: 1), TomL1 (SEQ ID NO: 2), NPTX2 (SEQ ID NO: 3) and ELN (SEQ ID NO: 4). In this way, information useful to determine the age of the individual is obtained. Optionally, for example, cytosine methylation is observed in both Edaradd
25 (SEQ ID NO: 1) and TomL1 (SEQ ID NO: 2); and/or both Edaradd (SEQ ID NO: 1) and NPTX2 (SEQ ID NO: 3). In some embodiments of the invention, cytosine methylation is observed in at least two genomic DNA sequences including ELN (SEQ ID NO: 4). Typically in these methods, cytosine methylation is observed one or more cytosine residues having a CG locus designation: cg09809672, cg27210390,

cg12799895 or cg05822532. Certain embodiments of the invention observe methylation in other gene sequences of the genes identified in Figures 4 and 5, such as TRIM58 as well as GRIA2 (see also Koch et al., (2011) AGING, Vol 3, No 10 , pp 1,018-1,027).

5 Related embodiments of the invention include methods of obtaining information useful to determine an age of an individual by obtaining a biological sample derived from an individual comprising genomic DNA from white blood cells or epithelial cells; and then observing a degree of cytosine methylation of at least one genomic DNA sequence of a gene selected from the group consisting of KCNG3,
 10 NPTX2, GREM1, VGF, GRIA2, PDE4C, FLJ42486, ATP8A2, KCNK12, C10orf82, ZIC1, BRUNOL6, ZNF667, TRIM58, ZNF540, B3GALT6, DCC, HTR7, IRF8, SKIP, CPLX2, POU4F2, KLF14, GATA4, TBX20, FLJ90650, NEFH, PCDH8, ADRB1, VMP, GLRA1, KCNC3, ZNF154, LEP, MCHR2, FAM19A4, RGC32, PCDHGB7, FBLN2, SLC15A3, PCDHGB4, SHOX2, LOC349136, KCNC1,
 15 TSPYL5, PNMA2, WNK2, ADRA2C, SLC5A7, BARHL2, LRRC2, NEF3, SPAG6, GCM2, IRXL1, NRN1, SMPD3, ZNF671, HOXB4, MOXD1, TCL1A, KCNA5, ASPA, Bles03, EDARADD, TCEA2, ELN, PGLYRP2, LGP1, TOM1L1, LAG3, SLC14A1, CSNK1D, ACSS2, C9orf66, CENTD3, HNRPL, CDH5, ABHD14A, LTBR, and RENBP. In embodiments of the invention, these sequences can include
 20 either translated or untranslated 5' regulatory regions; and optionally are within 1 kilobase (5' or 3') of the specific GC loci that are identified in Figure 4. These methods can further comprise correlating the degree/amount of cytosine methylation observed in the sequence with a predicted age of the individual, so that information useful to determine the age of the individual is obtained.

25 Embodiments of the invention include methods where observations of cytosine methylation in genomic DNA from a biological sample are used to predict the chronological age of the individual from which a sample is derived. In certain embodiments the methods of the invention can be used to provide valuable information in forensic investigations (e.g. where the identity of the individual from

which the DNA is derived is unknown). Such embodiments of the invention can be combined with other forensic analysis procedures, for example by also performing a DNA fingerprinting analysis on the genomic DNA. DNA fingerprinting (also known as DNA profiling) using short tandem repeats (STRs) is one method for human
5 identification in forensic sciences, finding applications in different circumstances such as determination of perpetrators of violent crime, resolving paternity, and identifying remains of missing persons or victims of mass disaster. The FBI and the forensic science community typically use 13 separate STR loci (the core CODIS loci) in routine forensic analysis. (CODIS refers to the Combined DNA Index System that
10 was established by the FBI in 1998). Illustrative DNA fingerprinting methodologies are disclosed, for example, in U.S. Patent Nos. 7,501,253, 7,238,486, 6,929,914, 6,251,592, and 5,576,180).

In addition, a measurement of relevant methylation patterns in genomic DNA from white blood cells or skin cells also provides a tool in routine medical screening
15 to predict the risk of age-related diseases as well as to tailor interventions based on the epigenetic bio-age instead of the chronological age. In some embodiments of the invention, one can compare the predicted age of the individual with the actual chronological age of the individual, for example as part of a diagnostic procedure for an age associated pathology (e.g. one that compares an individual's chronological age
20 with an apparent biological age in view of their DNA methylation patterns). Such methods can be useful in clinical interventions that are predicated on an epigenetic bio-age rather than an actual chronological age. In one embodiment, a biological sample can be collected in a routine health check and sent to the lab for methylation pattern analysis (e.g. as described above). If the predicted age of the patient is higher
25 than the real age, the patient can be at an increased risk of age-related diseases, and dietary intervention, or specific drugs, could be prescribed to reduce this "genetic age".

As noted above, embodiments of the invention include methods of obtaining information useful to determine a level of risk of an age-related disease in an

individual (e.g. Alzheimer's disease or Parkinson's disease). Typically, these methods comprise the steps of obtaining a biological sample derived from an individual comprising genomic DNA from white blood cells or epithelial cells and then observing a degree of cytosine methylation of at least one genomic DNA

5 sequence of a gene selected from the group consisting of KCNG3, NPTX2, GREM1, VGF, GRIA2, PDE4C, FLJ42486, ATP8A2, KCNK12, C10orf82, ZIC1, BRUNOL6, ZNF667, TRIM58, ZNF540, B3GALT6, DCC, HTR7, IRF8, SKIP, CPLX2, POU4F2, KLF14, GATA4, TBX20, FLJ90650, NEFH, PCDH8, ADRB1, VMP, GLRA1, KCNC3, ZNF154, LEP, MCHR2, FAM19A4, RGC32, PCDHGB7, FBLN2,

10 SLC15A3, PCDHGB4, SHOX2, LOC349136, KCNC1, TSPYL5, PNMA2, WNK2, ADRA2C, SLC5A7, BARHL2, LRRC2, NEF3, SPAG6, GCM2, IRXL1, NRN1, SMPD3, ZNF671, HOXB4, MOXD1, TCL1A, KCNA5, ASPA, Bles03, EDARADD, TCEA2, ELN, PGLYRP2, LGP1, TOM1L1, LAG3, SLC14A1, CSNK1D, ACSS2, C9orf66, CENTD3, HNRPL, CDH5, ABHD14A, LTBR, and RENBP. In

15 embodiments of the invention, these sequences can include either translated or untranslated 5' regulatory regions; and optionally are within 1 kilobase (5' or 3') of the specific GC loci that are identified in Figure 4. Embodiments of these methods further comprise calculating a theoretical bio-age of the individual based on the degree/amount of cytosine methylation observed in the sequence and then comparing

20 the theoretical bio-age of the individual to an actual chronological age of the individual. In this way, information useful to determine a level of risk of an age-related disease in the individual is obtained. Optionally for example, the theoretical bio-age of the individual is compared to an actual chronological age to determine if the theoretical bio-age is greater than the actual chronological age; and the method

25 further include providing an individualized treatment to the individual to bring the theoretical bio-age closer to the actual chronological age of the individual.

As noted above, embodiments of the invention can comprise correlating patterns of cytosine methylation with a predicted chronological age of the individual. For example, as disclosed herein, the methylation of certain sites such as the

promoters of the EDARADD, TOM1L1, and NPTX2 genes increases with age in a linear manner over a range of five decades. In this context, illustrative working embodiments of the invention use a regression analysis to correlate an observed pattern of cytosine methylation the predicted age of the individual. Using just two
5 cytosines from the EDARADD, TOM1L1, and NPTX2 loci, a regression model was designed that explains 73% of the variance in age, and is able to predict the age of an individual with an average accuracy of 5.2 years.

Using conventional regression model/analysis tools and methodologies known in the art, a number of age prediction models can be designed for use with specific
10 genomic DNA samples and/or specific analysis techniques and/or specific individual populations (see, e.g., statistical package R version 2.11.1 in citation as discussed in R Development Core Team (2005) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL www.R-project.org). In addition to the illustrative models
15 disclosed herein, other models can, for example, customize the coefficient values (weights) for different tissues and/or cell lineages. In addition to tissue type, such coefficients can be weighted in data sets from different populations. For example, if a model is applied to pediatric patients only, then one set of coefficients can be used. Alternatively, if a model is applied exclusively to older people (say >50 years),
20 another set of coefficients can be used. Alternatively, coefficients can be fixed, when for example, a model is broadly applied to people of ages from 10 to 100 etc. Coefficient values in various models can also reflect the specific assay that is used to measure the methylation levels (e.g. as the variance of the methylation levels of individual probes may affect the coefficient). For example, for beta values measured
25 on the Illumina methylation microarray platforms there can be one set of coefficients, while for other methylation measures (e.g. using sequencing technology) there can be another set of coefficients etc. In practicing certain embodiments of the invention, one can collect a reference data set (e.g. of 100 individuals of varying ages) using specific technology platform(s) and tissue(s) and then design a specific multivariate

linear model is fit to this reference data set to estimate the coefficients (e.g. using least squares regression). The resultant multivariate models can then be used for predicting ages on test patients. In this way, different mathematical models can be adapted for analyzing methylation patterns in a wide variety of contexts.

5 As discussed in detail below, in addition to using art accepted modeling techniques (e.g. regression analyses), embodiments of the invention can include a variety of art accepted technical processes. For example, in certain embodiments of the invention, a bisulfite conversion process is performed so that cytosine residues in the genomic DNA are transformed to uracil, while 5-methylcytosine residues in the
10 genomic DNA are not transformed to uracil. Kits for DNA bisulfite modification are commercially available from, for example, Human Genetic Signatures' Methyleasy and Chemicon's CpGenome Modification Kit. See also, WO04096825A1, which describes bisulfite modification methods and Olek et al. Nuc. Acids Res. 24:5064-6 (1994), which discloses methods of performing bisulfite treatment and subsequent
15 amplification. Bisulfite treatment allows the methylation status of cytosines to be detected by a variety of methods. For example, any method that may be used to detect a SNP may be used, for examples, see Syvanen, Nature Rev. Gen. 2:930-942 (2001). Methods such as single base extension (SBE) may be used or hybridization of sequence specific probes similar to allele specific hybridization methods. In another
20 aspect the Molecular Inversion Probe (MIP) assay may be used.

 In certain embodiment of the invention, the genomic DNA is hybridized to a complimentary sequence (e.g. a synthetic polynucleotide sequence) that is coupled to a matrix (e.g. one disposed within a microarray). Optionally, the genomic DNA is transformed from its natural state via amplification by a polymerase chain reaction
25 process. For example, prior to or concurrent with hybridization to an array, the sample may be amplified by a variety of mechanisms, some of which may employ PCR. See, for example, PCR Technology: Principles and Applications for DNA Amplification (Ed. H. A. Erlich, Freeman Press, NY, N.Y., 1992); PCR Protocols: A Guide to Methods and Applications (Eds. Innis, et al., Academic Press, San Diego,

Calif., 1990); Mattila et al., Nucleic Acids Res. 19, 4967 (1991); Eckert et al., PCR Methods and Applications 1, 17 (1991); PCR (Eds. McPherson et al., IRL Press, Oxford); and U.S. Pat. Nos. 4,683,202, 4,683,195, 4,800,159, 4,965,188, and 5,333,675. The sample may be amplified on the array. See, for example, U.S. Pat. No. 6,300,070 which is incorporated herein by reference.

Another embodiment of the invention is a kit for obtaining information useful to determine the age of an individual, the kit comprising a plurality of primers or probes specific for at least one genomic DNA sequence in a biological sample, wherein the genomic DNA sequences comprises a CG loci identified in Figure 4. Such kits of the invention can further include a reagent used in a genomic DNA polymerization process, a genomic DNA hybridization process, and/or a genomic DNA bisulfite conversion process. In some embodiments of the invention, the kit comprises a methylation microarray (see, e.g. U.S. Patent Application Publication no. 20060292585, the contents of which are incorporated by reference) . Optionally, the kit comprises a plurality of primer sets for amplifying at least two genomic DNA sequences. In certain embodiments of the invention, the kit comprises primer sets for amplifying at least two genomic DNA sequences in the biological sample selected from the group consisting of Edaradd (SEQ ID NO: 1), TomL1 (SEQ ID NO: 2), NPTX2 (SEQ ID NO: 3) and ELN (SEQ ID NO: 4). In some embodiments of the invention, the kit further comprises a probe or primer used to perform a DNA fingerprinting analysis.

EXAMPLES

The present invention is described in detail in the following example, but is not limited by any aspect of this example.

EXAMPLE 1: TYPICAL METHODS AND MATERIALS USEFUL FOR PRACTICING EMBODIMENTS OF THE INVENTION

Genomic DNA Collection Protocols

Saliva was collected using Oragene DNA collection kits (Genotek). The majority (up to 74%) of the DNA in saliva collected with this method comes from white blood cells, with the remainder being buccal epithelial cells (see, e.g. Thiede C, et al. (2000) Buccal swabs but not mouthwash samples can be used to obtain pretransplant DNA fingerprints from recipients of allogeneic bone marrow transplants. *Bone Marrow Transplant* 25: 575-577). Genomic DNA was prepared according to the manufacturer's protocol. Zygosity was determined using 9 microsatellite markers. Microarray hybridization was performed by the Southern California Genotyping Consortium at UCLA. 500 ng of genomic DNA was bisulfite converted using the EZ-methylation kit (Zymo Research), and processed according to the Illumina Infinium whole genome genotyping protocol. Labelled samples were hybridized to Illumina HumanMethylation27 arrays, scanned (iScan reader, Illumina), and beta (methylation) values extracted using GenomeStudio software.

Analysis: A signed weighted correlation network was constructed as described (see, e.g. Zhang B, et al. (2005) A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol* 4: Article17; Langfelder P, et al. (2008) WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 9: 559). Module definition was based on the gene methylation status in saliva and ignored age. As module representative, we used the module eigenlocus (ME) which is defined as the first principal component of the module methylation profiles and can be considered a weighted average. To incorporate age into the network analysis, the Student t-test statistic for correlating age with methylation status was used. Lasso penalized regression was performed using the 'penalized' package of R (see, e.g. Wu TT, et al. (2009) Genome-wide association analysis by lasso penalized logistic regression. *Bioinformatics* 25: 714-721). All statistical analyses and data processing were performed using the statistical package R version 2.11.1 (see, e.g. R Development Core Team (2005) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN

3-900051-07-0, URL <http://www.R-project.org>). PCR primers and conditions for amplification, massarray and pyrosequencing analysis are described below.

PCR protocol and primers

5 500 ng of genomic DNA was bisulfite converted using the EZ-methylation kit (Zymo Research), and 10 ng was used for each reaction. Edaradd was amplified using Bio-x-act Long (Bioline) with 20 seconds annealing at 57°C, and 30 seconds extension at 68°C, for 40 cycles. Tom1L1 and NPTX2 were amplified using Sahara Mix (Bioline) with 20 seconds annealing at 58°C (Tom1L1) or at 55°C (NPTX2), and
10 25 seconds extension at 72, for 40 cycles. Primer sequences for the amplicons are listed below. The appropriate sequence tags were added to the Edaradd PCR primers according to the instructions provided by Sequenom.

Edaradd methylation was assayed using MassArray (Sequenom) by the Genomics Core of the Albert Einstein College of Medicine, and the percent
15 methylation at each CpG site extracted with EpiTyper software (Sequenom). Tom1L1 and NPTX2 methylation was assayed using Pyrosequencing by the UCLA Genotyping and Sequencing Core.

Edaradd:

20 F: GGTAGATTAAGAGGAAGTTTATTTTTTTTAT (SEQ ID NO: 5)
 R: AATACCTCTCCCCATCTATTTAATC (SEQ ID NO: 6)

Tom1L1:

 F: TTAATTTATTGTAGAATTTT (SEQ ID NO: 7)
25 R: AAACCTCCTCTTCTAATCTATAAAAC (SEQ ID NO: 8)

Sequencing primer: ATAAAATATTTAAACCTCCA (SEQ ID NO: 9)

NPTX2:

F: TAGTTTAAGAAAGGG (SEQ ID NO: 10)

R: AACTATCCTAAACCCCAAC (SEQ ID NO: 11)

Sequencing primer: ACAAAAACTTCTACCC (SEQ ID NO: 12)

5

Microarray analysis

Monozygotic (MZ) twins form an attractive model to study methylation changes with age. At the time of separation both embryos have nearly identical methylation patterns. While certain methylation changes are genetically controlled, 10 environmental exposure and stochastic processes can also lead to a change in methylation patterns. In this context, identical twins can be considered replicates of the same developmental and aging experiment.

In this study we quantified the methylation status of 27,578 CpG loci covering more than 14,000 genes at single-nucleotide resolution in saliva samples of 34 pairs of 15 identical twins, between 21 and 55 years of age, using Illumina HumanMethylation27 microarrays. Monozygosity was verified for all pairs by analysis of nine short tandem repeat probes. For each CpG site on the microarray, we calculated the beta value, which expresses the fraction of methylated cytosines in that location. A site that is completely methylated on both alleles in all cells has a beta value equal to 1; a 20 completely unmethylated site equals 0. All subsequent analyses were performed on this beta value. For computational reasons, the data was filtered by requiring a mean methylation value between 0.05 and 0.95, and variance greater than 0. The resulting restricted dataset contained 16,155 probes, and all further analyses were performed on this filtered dataset. Batch effect were removed using the Combat algorithm (see, e.g. 25 Johnson WE, et al. (2007) Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* 8: 118-127), and one outlier sample was removed.

We first determined whether methylation differences measured using these arrays reflected actual differences between individuals by calculating the correlation

coefficient between replicate arrays for 10 samples. The median correlation between replicate arrays was 0.995 (range 0.990-0.996), compared to 0.987 (range 0.957-0.994) between unrelated samples. This difference was highly significant (Wilcoxon test, $p=1.4 \times 10^{-7}$). In unsupervised hierarchical clustering, the majority of twin pairs
5 clustered together (Figure 7) and twin samples correlated with $r=0.992$ (range 0.983-0.997), which is significantly different from the correlation between unrelated samples (Wilcoxon test, $p=1.93 \times 10^{-11}$).

A previous study showed increasing global epigenetic differences with age in a sample of identical twins, suggesting increased epigenetic drift with age (see, e.g.
10 Fraga MF, et al. (2005) Epigenetic differences arise during the lifetime of monozygotic twins. *Proc Natl Acad Sci USA* 102: 10604-10609). We were unable to replicate these genome-wide methylation changes when the intra-pair correlation coefficients, the intra-pair Euclidian distance, or the intra-pair Manhattan distance was correlated with age ($p>0.1$). We did, however, identify a subset of loci to be highly
15 correlated with age.

A recurrent problem with data analysis on a whole genome scale is correcting for multiple comparisons. The stringency level of the chosen correction method strongly affects the odds of identifying significant findings. We previously described weighted correlation network analysis (WGCNA) as a data reduction scheme (see,
20 e.g. Horvath S, et al. (2008) Geometric interpretation of gene coexpression network analysis. *PLoS Comput Biol* 4: e1000117; Zhang B, et al. (2005) A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol* 4: Article17). Here we used WGCNA to identify modules of loci with highly similar methylation values. First, we averaged all methylation values for each twin pair, and
25 treated each pair's data as an individual sample. Since both twins are genetically identical and of the same age, averaging the data reduces possible environmental effects on DNA methylation. After hierarchical clustering of the data set, branches of the cluster dendrogram defined five modules ranging in size from 199 to 842 loci, of which the methylation values were highly correlated across the samples (Figure 1A).

We color-coded the modules, calculated a weighted average, representative locus (eigenlocus) for each module (see methods) and correlated this with age. The correlation between age and the representative of the green module was highly significant ($r=0.62$, $p=7.2 \times 10^{-5}$, Figure 1B), even after using the most stringent multiple comparison correction (Bonferroni), since only 5 comparisons—corresponding to 5 modules—were carried out. Module membership of all probes can be found in Figure 7.

Identification of 88 novel loci correlated with age

To identify novel loci for which the methylation values correlate positively or negatively with age, we calculated q -values to correct for multiple comparisons (see, e.g. Storey JD, et al. (2003) Statistical significance for genomewide studies. *Proc Natl Acad Sci USA* 100: 9440-9445). We selected probes with q -values smaller than 0.05, corresponding to absolute correlation values greater than 0.57. A total of 88 probes correlated with age (Figure 4), corresponding to 80 genes spread over several of the modules. Of these, 19 probes were negatively correlated, and 69 were positively correlated with age, of which 57 belonged to the green module. A recently published study used a very similar study design and identical microarrays to identify 131 CpG sites correlated with age in blood samples of identical twins ranging from 49 to 75 years of age (see, e.g. Rakyan VK, et al. (2010) Human aging-associated DNA hypermethylation occurs preferentially at bivalent chromatin domains. *Genome Res* 20: 434-439). Of these 131 sites, 10 were found to be positively correlated with age in our study as well (Figure 5).

Of the 88 probes that were significantly correlated with age in our study, only one was near a gene encoding a microRNA (HSA-MIR-10A, in the HOXB4 gene), which was not different from the density on the array. 73 of 88 (83%) significant probes were within CpG-islands, thus this probe set was enriched in CpG islands relative to the typical array probe (73% in CpG islands, $p=0.031$, Fisher's exact test

for count data). CpG sites that were significantly correlated with age were a median 238 base-pairs upstream of the transcription start site.

Ingenuity analysis showed the 80 age-related genes were highly enriched for genes involved in cardiovascular disease ($p=1.59 \times 10^{-6}$), neurological disease
 5 ($p=1.47 \times 10^{-4}$), and genetic disease ($p=1.59 \times 10^{-6}$)—a category consisting almost entirely of the cardiac and neurological genes as well. The most enriched cellular function was molecular transport ($p=2.4 \times 10^{-3}$). The full gene ontology analysis can be found in Figure 6.

10 Validation of correlated probes in additional samples

Three probes for which the methylation status was highly correlated with age, and which had the widest distribution of values, were chosen for further validation. Saliva samples from 22 twins from the array study, 31 unrelated male, and 29 unrelated female samples (age range = 18-70 years-old) were bisulfite converted and
 15 PCR amplified. The fraction of methylated cytosines at the exact CpG sites assayed on the Illumina arrays were quantified by MassArray (Sequenom) for the Edaradd gene and by pyrosequencing for NPTX2 and Tom1L1. For NPTX2, the pyrosequencing method provided methylation data for five additional CpG sites in the promoter. The results of the validation experiments correlated very strongly with the
 20 array data for all three genes (Edaradd $r=0.96$, NPTX2 $r=0.92$, Tom1L1 $r=0.90$, $n=23$), providing a technical replication of the array data in the twin sample. The correlation between the degree of methylation and age of all three genes was preserved in the subset of twins and was also found in the independent male sample, providing a biological replication. In females, Edaradd and Tom1L1 were
 25 significantly correlated with age, but NPTX2 was not. The correlation results are shown in Figure 2. A multivariate linear regression model using Edaradd, Edaradd squared and NPTX2 showed that these two markers explain 76% (or $R^2=0.76$) of the variance in age of males and 70% in females. When considering males and females together the model explained 73% of the variance in age.

A leave-one-out analysis forms an accurate epigenetic predictor of age

To provide an unbiased estimate of predictive accuracy for age, we used a leave-one-out analysis where the multivariate regression model was fit on all but one subject and its prediction was related to the truly observed age of the left-out subject.

- 5 The predicted values are highly correlated with the observed age in males ($r=0.83$, $p=3.3 \times 10^{-13}$, $n=47$, Figure 8), females ($r=0.75$, $p=2.4 \times 10^{-4}$, $n=19$, Figure 9), and in the combined sample ($r=0.83$, $p=2.2 \times 10^{-16}$, $n=66$, Figure 3). For the male only or female only models, the average absolute differences between the predicted and the observed age (the error) are 5.3 years and 6.2 years, and for the combined sample this
- 10 is 5.2 years. Even when only the male and female replication samples were used, discarding all twin data, the accuracy of the model remained at 5.3 years, and the predicted values correlated highly with the observed age ($r=0.85$, $p=1.701 \times 10^{-13}$, $n=45$, Figure 10).

- To test whether additional data points on the microarray could improve the accuracy of the model, we performed lasso penalized regression to screen for the top
- 15 predictors of age (see, e.g. Tibshirani R (1997) The lasso method for variable selection in the Cox model. *Stat Med* 16: 385-395; Wu TT, et al. (2009) Genome-wide association analysis by lasso penalized logistic regression. *Bioinformatics* 25: 714-721). The top five predictors were tested, and only three were found to
- 20 contribute significantly to the regression model: Edaradd, NPTX2, and ELN. The first two predictors were already part of the model. Using the microarray methylation data for these two genes, the average error is 4.7 years ($r=0.77$, $p=1.029 \times 10^{-7}$, $n=34$). Adding the ELN methylation data improved the accuracy of our model, reducing the average error to 3.5 years ($r=0.87$, $p=2.2 \times 10^{-11}$, $n=34$, Figure S5).
- 25 Results were nearly identical when all twin samples were treated as unrelated individuals, and when averaged values for each pair were used. The distribution of methylation values for ELN was considered too narrow for further validation using pyrosequencing or MassArray analysis.

DISCUSSION AND CHARACTERIZATION OF EXPERIMENTAL DATA

In this high density, genome-wide screening of CpG methylation of twins, we identified 88 CpG sites near 80 genes for which the percent methylation in saliva is significantly correlated with age. These are highly enriched for genes known to influence age-related diseases—mainly cardiovascular and neurological disease. Ten of these 88 CpG sites were shown earlier to be correlated with age in whole blood and in isolated CD4⁺ and CD14⁺ cells as well (see, e.g. Rakyan VK, et al. (2010) Human aging-associated DNA hypermethylation occurs preferentially at bivalent chromatin domains. *Genome Res* 20: 434-439). We validated three genes in a sample of unrelated males and females, which confirmed our findings in these replicate samples. Remarkably, the methylation values for the validated genes are linear with age over a span of five decades and in three separate sample sets. Based on this observation, we were able to build a model that can predict the age of a subject based on the methylation status of just two cytosines in the genome, explaining 73% of the variance in age.

Of the validated genes, Neuronal Pentraxin II (NPTX2) methylation has been shown to be upregulated in pancreatic cancer (see, e.g. Park JK, et al. (2007) Quantitative analysis of NPTX2 hypermethylation is a promising molecular diagnostic marker for pancreatic cancer. *Pancreas* 35: e9-15), and its expression is increased in Parkinson's disease (see, e.g. Moran LB, et al. (2008) Neuronal pentraxin II is highly upregulated in Parkinson's disease and a novel component of Lewy bodies. *Acta Neuropathol* 115: 471-478). Its methylation status was recently shown to be correlated with age in blood as well (see, e.g. Rakyan VK, et al. (2010) Human aging-associated DNA hypermethylation occurs preferentially at bivalent chromatin domains. *Genome Res* 20: 434-439). Mutations in the Edar associated death domain (Edaradd) can cause loss of hair, sweat glands, and teeth (see, e.g. Yan M, et al. (2002) Identification of a novel death domain-containing adaptor molecule for ectodysplasin-A receptor that is mutated in crinkled mice. *Curr Biol* 12: 409-413), and it can reduce the speed of wound healing (see, e.g. Langton AK, et al. (2008) An

extended epidermal response heals cutaneous wounds in the absence of a hair follicle stem cell contribution. *J Invest Dermatol* 128: 1311-1318). Further research should focus on their role in aging, and age-related diseases.

The lack of epigenetic drift within each monozygotic pair contrasts with a previous study (see, e.g. Fraga MF, et al. (2005) Epigenetic differences arise during the lifetime of monozygotic twins. *Proc Natl Acad Sci USA* 102: 10604-10609). The main difference between the two studies is that we focused on CpG sites close to functional gene transcription start sites whereas Fraga and colleagues investigated random sites, most of which were located in non-functional repeated sequences (e.g., Alu repeats). This suggests that while drift may occur randomly with age in non-coding, repeat-rich DNA regions, the critical regulatory portions of the genome remain under strict epigenetic control throughout life.

Our regression model (Figure 3) could be applied in a variety of contexts. For instance, our ability to predict an individual's age to an average accuracy of 5.2 years could be used by forensic scientists to estimate a person's age based on a biological sample alone, once the model has been tested in various biological tissues. The model is also relevant to healthcare applications. Previously, significant DNA methylation differences were shown to be associated with specific age-related disorders, for example in comparisons between the brains of people diagnosed with late-onset Alzheimer's disease and brains from controls (see, e.g. Wang SC, et al. (2008) Age-specific epigenetic drift in late-onset Alzheimer's disease. *PLoS One* 3: e2698). The identification of specific epigenetic patterns highly correlated with age has the potential to influence our understanding of aging in health and disease. Specifically, it could lead to clinical interventions that are tailored to patients based on their "bio-age"—a result of the interaction of genes, environment, and time—rather than their chronological age. Future investigations should focus on phenotype and disease history of those subjects whose predicted age vary widely from their actual age. Furthermore, these findings could pave the way for interventions based on specific epigenetic marks associated with disease, as is already the case in cancer treatment

(see, e.g. Marks PA, et al. (2007) Dimethyl sulfoxide to vorinostat: development of this histone deacetylase inhibitor as an anticancer drug. *Nat Biotechnol* 25: 84-90).

TABLE 1: ILLUSTRATIVE SEQUENCES IN THE Edaradd, Tom1L1, NPXT2 and ELN GENES

In this table, illustrative primer sequences are shown in **bold**. Note that the sequence shown is the genomic DNA sequence, while the primer sequences are designed to bind to bisulfite converted DNA. Hence the fact that the given primer sequences do not align exactly with the genomic sequences shown. Sequencing primers are in **bold and underlined**. Certain CpG loci correlated with age underlined. Additional CpGs that are assayed in the same experiment and can be correlated with age are in *italics*.

Edaradd:

15 F: **GGTAGATTAAGAGGAAGTTTATTTTTTTAT** (SEQ ID NO: 5)
 R: **AATACCTCTCCCCATCTATTTAATC** (SEQ ID NO: 6)

TTGTATGGGAACCTCTGGTGAATGCGAATCATTTTTAAATTACTTTTTTTGTAAAGTGCAAAACAACAAT
 AGCACCCATTTGCGTCATACTTTTATAGTTTCGCAAAGCACATGGGAAAAATAAAGGTAATGATGGGGATC
 GTTGCAATTCATAGGAAAGGAGGCACGAGGAAATGAAAATGAAAGGGAGTAATAACTACGTAAGTAGTC
 20 AATCTTCCTTAAAAAAAAAACCTTAAAAATATACCACCATCTTCTATTTGATATAATGCAGAATGGGA
 ATGATAAAACATGAATTACATTTTCAGAGTTTCAAAAAGCAAACCAGCTTTATAGCAATGCTTGAGGTT
 GGGCTGCTAACAAGCTCACTCAACTAGTGTTTCTGACGGCCAACGTCAGAATAATTCCATCTCCATGA
 GAAGTACAGAAAGAACCACAAACCAACCTCCAAATTGATTCTAAGATAAAATACCTTAAAAAAAATT
 TCCCTTCTATCCGGGC**GGCAGACCAAGAGGAAGTTTATCCTCCCACCT**TACAAATTCCCCAGAGAGCTT
 25 TCATCTAGAAGGTTTGACTCTGGCCAGACAACCAGCGAGCATCTTCTCGCAATCTGTTGCTTCTTCCAT
 GGCAAACTCCAGAGAATTAAGAAGCCAAACTCAACATCGCCATGGGCCTCAGGAC**GACTAAACAGATGG**
GGAGAGGCACTGGCAGACCAAGAGGAAGTTTATCCTCCCACCTACAAATTCCCCAGAGAGCTTTCATCT
 AGAAGGTTTGACTCTGGCCAGACAACCAGCGAGCATCTTCTCGCAATCTGTTGCTTCTTCCATGGCAAA
 CTCCAGAGAATTAAGAAGCCAAACTCAACATCGCCATGGGCCTCAGGACGACTAAACAGATGGGGAGAG
 30 GCACTAAAGCTCCTGGTCACCAAGAGGGTATGTAGGCATTTGCTGTCTTCTGGATTTCTCAGAGCTGA
 GTTTTGTAGCCAGAGGTGCTTATTTACGATAATTCTTGGATATATTATACACTAAATACTATTATTATC
 TTTTTCGACCCGACTTTTATCTTCTGTTCTTATGTGTGAAGGCAGAGAAAGATTATTTAGAGCTCTTC
 AAAGATTCTTATTTAATTTAAATGCCTGTGCTTCTTATAATAGGCTTATGATGGATGATAGCTTTA
 GTTAAATGTAGCAATCTTAAATATATT (SEQ ID NO: 1)

Tom1L1:F: **TTAATTTATTGTAGAATTTT** (SEQ ID NO: 7)R: **AAACCTCCTCTTCTAATCTATAAAAC** (SEQ ID NO: 8)5 Sequencing primer: **ATAAAATATTTAAACCTCCA** (SEQ ID NO: 9)

CTCCTCGCGGGTCCCGCAGGGCGCGCTGCAGGTGCGCTGGGCGTCCCGAAGCCCCGCCCTCGCGTCCCT
 GCCCCGCGCCAGCTCCTCACGCTCGGTGCCCCGCGCCCCGCAGCGGCAGCGCAGAACGCACCGCCTCTG
 CCAGAGCCCCGGAAGCGCTCGGGCGAAGAGGAGGAGCCAAGGGTACCGAGCGGGTGGAGTCGGGAGCCG
 GAGAGCGGTGGAGGCGGATTTCTTGGGCCCCGCCCTCTGGCGCTACCATGGCGTTTGGCAAGAGTCACC
 10 GGGATCCCTACGCGACCTCCGTGGGCCACCTCATAGGTAAGGAGGCGCGGGAGAGACGCCAGGCAGG
 CAGGGGACCGTGGGATCCTTTCTGCTTGATCCATTCTCGGCCTGCAGAGGACGGAGTTAGTCCAACCTT
 GAAAAAATTATTCCCTCCCCCGCAACTTTCCCAAGGCACCCGCATTCCACCCGGCCCCCTTTTCGTG
 TTTCTTCTTGG**CCAACTCACTGTAGAACCTC**AGTCCTCAAAAAATGTACCTTCCTTTTCGATGCCGCTG
 GGGAGTGGAACCAAACAGGTGAACCGCGGGAGTCAGGC**ATGGAGTGTTGGGCCTCCA**CGAGGAGACAC
 15 CAGAACTTCTCGGTAGGGGAAGTTATTCTTAAAGGCACATTCTCCAGGGCACGGTATTGTTATGCCCCG
TTTTACAGATCAGAAGAGGAGGCTCTTGGGAGACTAAGTGATGTGGCCTTACTGTGAGGGGCAGGTCTCC
 CCAAACCAGTTTTCTCTAGTGCCAAGCTTATACAACCTGGATTACTTTTCTGGGTAAACAAAGAGAATG
 AAAATGCTTGGCATTGGAGTGAAGAGGAGGGGAACATAGGAGCCATCTGGGGCCACGGAGGTCTCTCGC
 AGGCCACTAGGGACTACACCAGCATCCACCTAAATGATTTCTCTCTCTCCGGATGTACCTTGCATGAT
 20 GTATGTAGAAGCCAGTGATTGGACTCCTACCAGCACCAGCTCACTGCTCCTTCTTTTCCCTGATTCTC
 CCTTGGGAATTGGGTGTGGCACGAGTGCATATTAATGCCAGGCAAAGTGTGATGATGTCAGGAACCTTAG
 TGGGGTGATGAAATCGTGAGAACCTAGTAGTTGTAGTAAAGAGGCCAGATTGGAAAATTTAGATCAC
 AGCTCTACTAACTGCTGTGACACCGAACACACCACTTAAACAGAACCT (SEQ ID NO: 2)

25 **NPTX2:**F: **TAGTTTAAGAAAGGG** (SEQ ID NO: 10)R: **AACTATCCTAAACCCCAAC** (SEQ ID NO: 11)Sequencing primer: **ACAAAAAATTCTACCC** (SEQ ID NO: 12)

CTCCTTCCCGCCTCGAGAGTGAGGTGGCCGGGCGCTTGACGAGAAGGCCACGCCTGCCGCGGGGGTGGC
 30 TCGCGATGGCAGTCGGGGTTCGAGTCCCGCCTGGGGGGCTGCTCCTGCTGGAGAAAACGCCTCCCTGAG
 GCGGCGGCAAACGCGCAGCGAGGCCCCGTGCCGCGCCAGAAGCCACCCTGAGAAAGGGGCACCGGGAC
 ACCGAGGGGTTCCTACTTTCTCCTCAGCCTGTGACGCCGCGTCTCGGGTGGGTTCGAGGGGCGCCTG
 GGCACGGCCAGCCGAGGCTCTCGAGAGCCCCAGTGTGTTTTCCACCTCAGGCCTCCTTTCTGAGGCA
 GAGCCCCGGACCTCGCGCTCTCGCCTCAGGCTCCGGCCACGCTCCCGCCCGCCGCGCCAGGCGCGCAAC
 35 GGAAAGCGCCCCCGCCCCGCTCCGCCCCACTGCGTGACGCGCACCCGGCCGAGCCAATCAGAGCT

CGTGGCGCGCGCCCCACACGCCGGCCCCCTCCGCCCT**CAGCTTAAGAAAGGG**CGCGCGGACCCGGCAG
 GCCAGAGTGCCGAGCAGCGCGGTGGGTGCGGCTGTGAGACG**GCAGGAGACTTCTGCCCCGCGGTGCACG**
CGACCCCTCGAGACGACAGCGCGGCTACTGCCAGCAGCGAAGGCGCCTCCCGCGGAGCGCCCCGACGGCG
 CCCGCTCGCCCATGCGCAGCTGAGCGCGGCAGCGGCGGCGGGATGCTGGCGCTGCTGGCCGCCAGCGTG
 5 GCGCTCGCGCTGGCC**GCTGGGGCCCAGGACAGCC**CGGCGCCCCGGTAGCCGCTTCGTGTGCACGGCACTG
 CCCCCAGAGGCGGTGCACGCCGGCTGCCCCGCTGCCCCGCGATGCCCATGCAGGGCGGCGCGCAGAGTCCC
 GAGGAGGAGCTGAGGGCCGCGGTGCTGCAGCTGCGCGAGACCGTCGTGCAGCAGAAGGAGACGCTGGGC
 GCGCAGCGCGAGGCCATCCGCGAGCTCACGGGCAAGCTAGCGCGCTGCGAGGGGCTGGCGGGCGGCAAG
 GCGCGCGGCGCGGGGGCCACGGGCAAGGACACTATGGGCGACCTGCCGCGGGACCCCGGCCACGTCGTG
 10 GAGCAGCTCAGCCGCTCGCTGCAGACCCCTCAAGGACCGCCTGGAGAGCCTCGAGGTAGCGGCCCGCGGG
 GAGCGCGGGGGACCTGGAATGGGGACGCTCCCAGCTCGGGGGCGGAAG (SEQ ID NO: 3)

ELN

The locus measured for the Elastin (ELN) gene is cg05822532. The actual CpG site
 15 measured is in underlined here:

CCTCCCTCTTTCCCTCACAGCCGACGAGGCAACAATTAGGCTTTGGGGATAAAACGAGGTGCGGAGAGC
 GGGCTGGGGCATTCTCCCCGAGATGGCGGGTCTGAC**GGCGGCGGCCCCGCGGCCCGGAGTCCTCCTGC**
 TCCTGCTGTCCATCCTCCACCCCTCTCGGCCTGGAGGTAAGGACCCCTCGCCCCGTCCCCAGCGCTGC
 CCACA (SEQ ID NO: 4)

20

The exact location in the March 2006 build of the human genome (NCBI36/hg18) is
 chromosome 7, pos 73080467.

This concludes the description of the preferred embodiment of the present
 25 invention. The foregoing description of one or more embodiments of the invention
 has been presented for the purposes of illustration and description. It is not intended to
 be exhaustive or to limit the invention to the precise form disclosed. Many
 modifications and variations are possible in light of the above teaching. It is intended
 that the scope of the invention be limited not by this detailed description, but rather by
 30 the claims appended hereto.

CLAIMS

1. A method of obtaining information useful to determine an age of an individual, the method comprising the steps of:
 - 5 (a) obtaining a biological sample derived from an individual comprising genomic DNA from white blood cells or epithelial cells; and
 - (b) observing a pattern of cytosine methylation of at least one genomic DNA sequence selected from the group consisting of Edaradd (SEQ ID NO: 1), TomL1 (SEQ ID NO: 2), NPTX2 (SEQ ID NO: 3) and ELN (SEQ ID NO: 4);
- 10 so that information useful to determine the age of the individual is obtained.
2. The method of claim 1, wherein cytosine methylation is observed in:
 - (a) Edaradd (SEQ ID NO: 1) and TomL1 (SEQ ID NO: 2); and/or
 - (b) Edaradd (SEQ ID NO: 1) and NPTX2 (SEQ ID NO: 3).
- 15 3. The method of claim 1, wherein cytosine methylation is observed in at least two genomic DNA sequences including ELN (SEQ ID NO: 4).
4. The method of claim 1, wherein a bisulfite conversion process is performed so
20 that cytosine residues in the genomic DNA are transformed to uracil, while 5-methylcytosine residues in the genomic DNA are not transformed to uracil.
5. The method of claim 1, wherein the genomic DNA is hybridized to a complimentary sequence disposed on a microarray.
- 25 6. The method of claim 1, wherein the genomic DNA is amplified by a polymerase chain reaction process.

7. The method of claim 1, wherein cytosine methylation is observed one or more cytosine residues having a CG locus designation: cg09809672, cg27210390, cg12799895 or cg05822532.
- 5 8. The method of claim 1, wherein the individual's identity is unknown.
9. The method of claim 1, further comprising performing a DNA fingerprinting analysis on the genomic DNA.
- 10 10. The method of claim 1, further comprising using the information in a forensic investigation.
11. The method of claim 1, further comprising correlating the pattern of cytosine methylation observed with a predicted age of the individual.
- 15 12. The method of claim 1, further comprising comparing the predicted age of the individual with the actual chronological age of the individual, wherein the comparison comprises a diagnostic procedure for an age associated pathology.
- 20 13. The method of claim 11, wherein a regression analysis is used to correlate the pattern of cytosine methylation observed with the predicted age of the individual.
14. A method of obtaining information useful to determine an age of an individual, the method comprising the steps of:
- 25 (a) obtaining genomic DNA from white blood cells or epithelial cells derived from the individual; and
- (b) observing cytosine methylation of at least two CG loci in the genomic DNA selected from the group consisting of CG locus designation: cg27553955, cg12799895, cg21296230, cg04084157, cg25148589, cg17861230, cg03734874,

- cg18236477, cg27320127, cg07621046, cg14456683, cg21801378, cg00107187, cg05508084, cg07533148, cg03975694, cg19945840, cg00399483, cg06291867, cg24826867, cg06092815, cg19885761, cg24199834, cg04528819, cg13434842, cg02008154, cg25044651, cg02994956, cg20366906, cg27389185, cg14826456,
- 5 cg12111714, cg12457773, cg11981599, cg00059225, cg06572160, cg08668790, cg12782180, cg20134215, cg13921352, cg13614181, cg23563234, cg00201234, cg21992250, cg00911351, cg19594666, cg15425280, cg06156376, cg19831077, cg27409364, cg15747595, cg02154186, cg20616414, cg10235817, cg16232126, cg24646414, cg17241310, cg10031651, cg23290344, cg06908778, cg02844545,
- 10 cg14614211, cg25511429, cg15201635, cg19246110, cg06760035, cg13603171, cg13282837, cg20792062, cg02228185, cg13547237, cg09809672, cg01293143, cg05822532, cg07408456, cg08468689, cg27210390, cg01820374, cg17589341, cg19761273, cg03440846, cg08909157, cg11136562, cg16464322, cg08872742, cg18328933, cg15784615 and cg23282949;
- 15 so that information useful to determine the age of the individual is obtained.

15. The method of claim 14, further comprising correlating cytosine methylation observed in the CG loci with a predicted chronological age of the individual.

- 20 16. A kit for obtaining information useful to determine the age of an individual, the kit comprising:

A plurality of primers or probes specific for at least two genomic DNA sequences in a biological sample, wherein the genomic DNA sequences comprise CG loci in the genomic DNA selected from the group consisting of CG locus designation:

25 cg27553955, cg12799895, cg21296230, cg04084157, cg25148589, cg17861230, cg03734874, cg18236477, cg27320127, cg07621046, cg14456683, cg21801378, cg00107187, cg05508084, cg07533148, cg03975694, cg19945840, cg00399483, cg06291867, cg24826867, cg06092815, cg19885761, cg24199834, cg04528819, cg13434842, cg02008154, cg25044651, cg02994956, cg20366906, cg27389185,

cg14826456, cg12111714, cg12457773, cg11981599, cg00059225, cg06572160,
 cg08668790, cg12782180, cg20134215, cg13921352, cg13614181, cg23563234,
 cg00201234, cg21992250, cg00911351, cg19594666, cg15425280, cg06156376,
 cg19831077, cg27409364, cg15747595, cg02154186, cg20616414, cg10235817,
 5 cg16232126, cg24646414, cg17241310, cg10031651, cg23290344, cg06908778,
 cg02844545, cg14614211, cg25511429, cg15201635, cg19246110, cg06760035,
 cg13603171, cg13282837, cg20792062, cg02228185, cg13547237, cg09809672,
 cg01293143, cg05822532, cg07408456, cg08468689, cg27210390, cg01820374,
 cg17589341, cg19761273, cg03440846, cg08909157, cg11136562, cg16464322,
 10 cg08872742, cg18328933, cg15784615 and cg23282949; and

a reagent used in:

a genomic DNA polymerization process;

a genomic DNA hybridization process; or

a genomic DNA bisulfite conversion process.

15

17. The kit of claim 16, wherein the kit comprises a plurality of primer sets for amplifying at least two genomic DNA sequences.

18. The kit of claim 17, wherein the kit comprises primer sets for amplifying at
 20 least two genomic DNA sequences in the biological sample selected from the group consisting of Edaradd (SEQ ID NO: 1), TomL1 (SEQ ID NO: 2), NPTX2 (SEQ ID NO: 3) and ELN (SEQ ID NO: 4).

19. The kit of claim 16, wherein the kit further comprises a probe or primer used
 25 to perform a DNA fingerprinting analysis.

20. The kit of claim 16, wherein the kit comprises a methylation microarray.

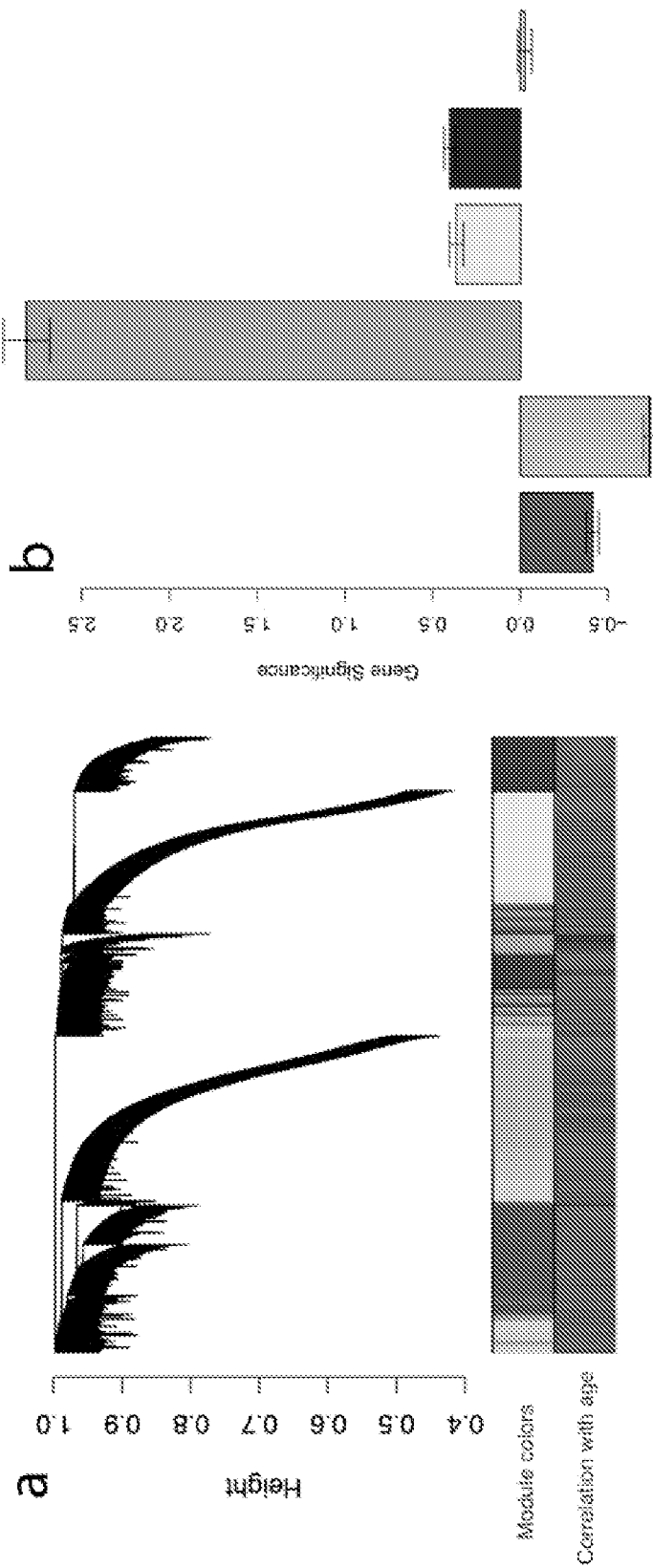
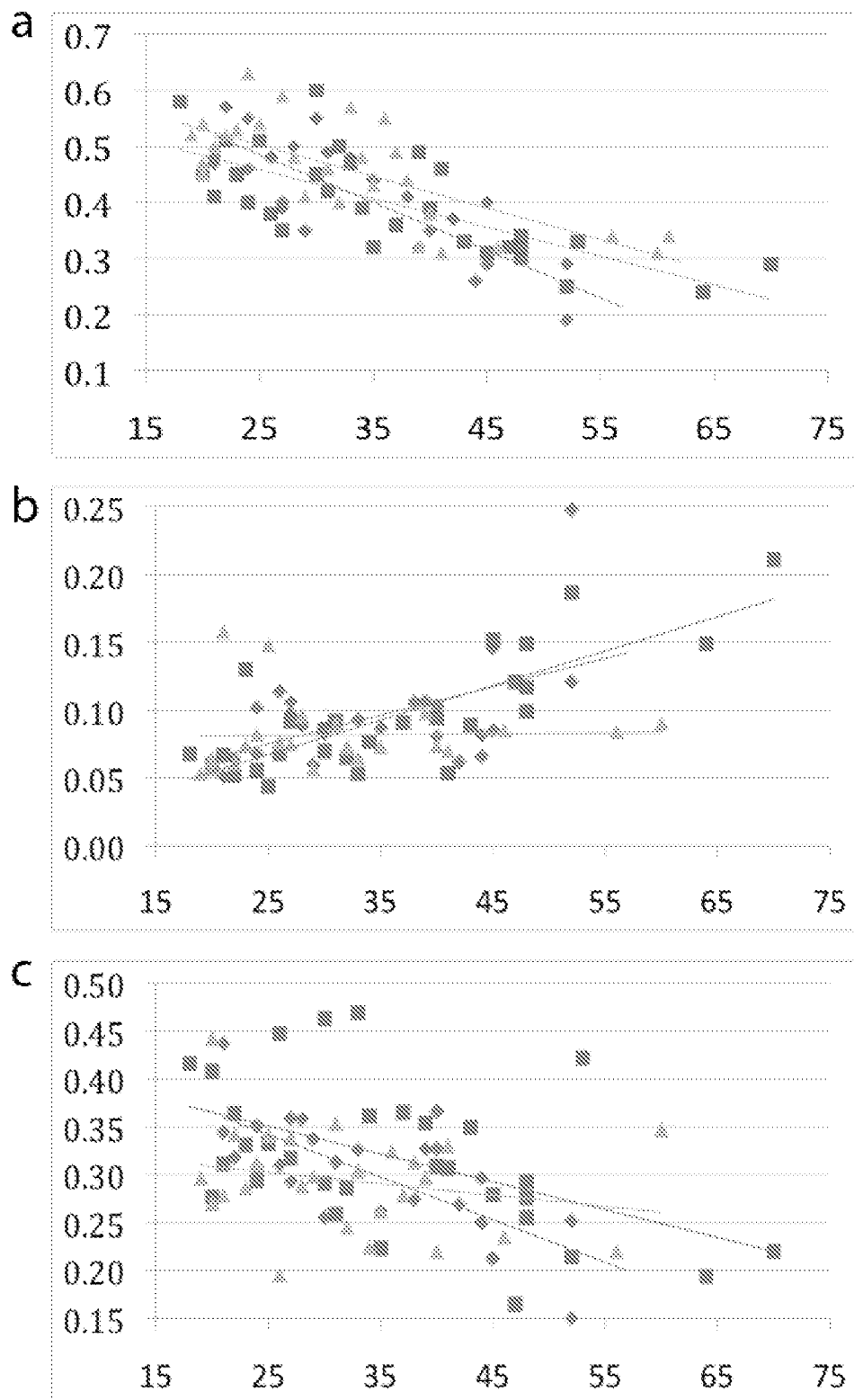


FIG. 1

**FIG. 2**

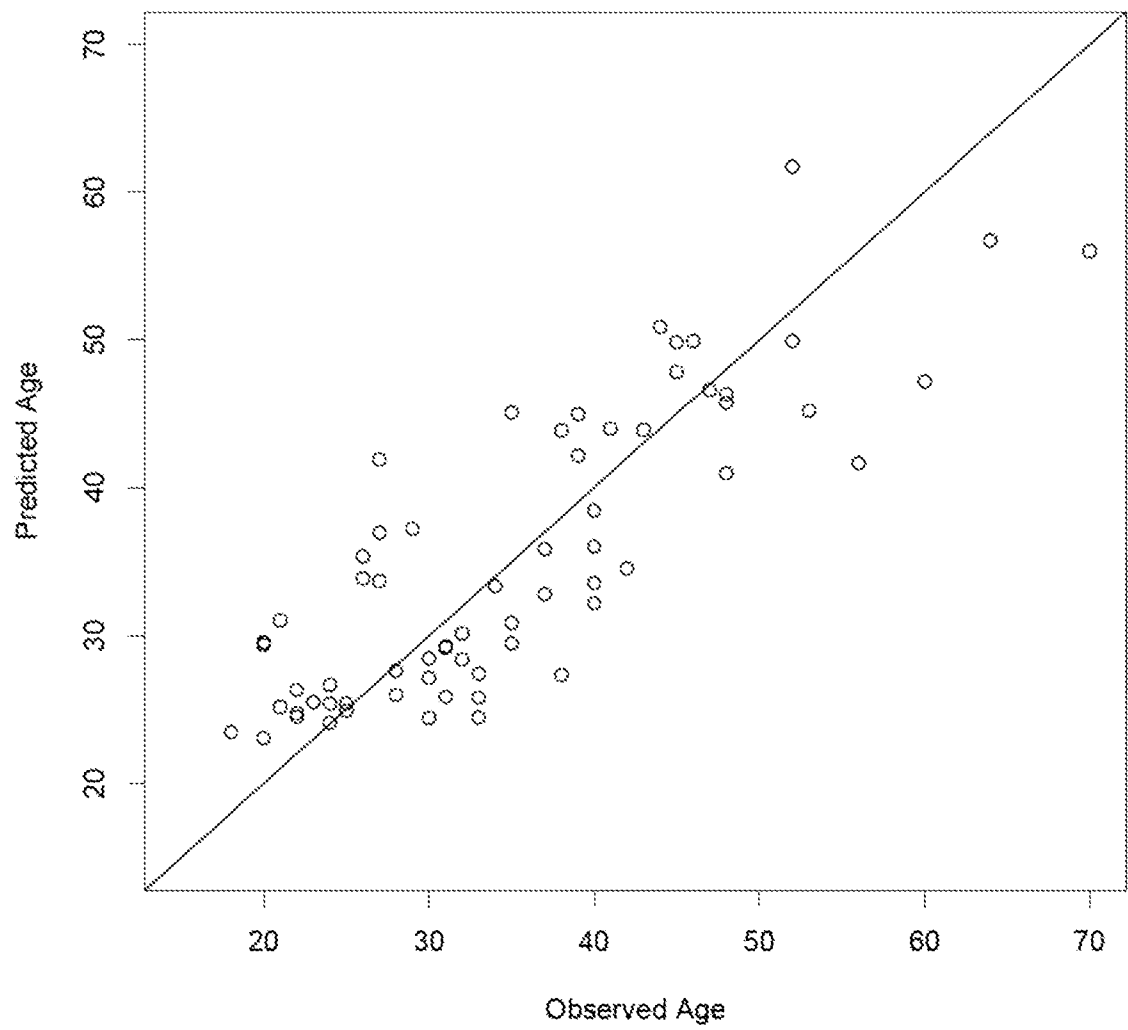


FIG.3

FIG. 4A

TargetID	Chr	Gene_ID	Symbol	r	p-value	q-value
cg27553955	2	170850	KCNG3	0.807	8.38E-09	0.00012
cg12799895	7	4885	NPTX2	0.754	2.59E-07	0.00124
cg21296230	15	26585	GREM1	0.758	2.14E-07	0.00124
cg04084157	7	7425	VGF	0.726	1.20E-06	0.00188
cg25148589	4	2891	GRIA2	0.723	1.37E-06	0.00188
cg17861230	19	5143	PDE4C	0.734	8.00E-07	0.00188
cg03734874	14	388021	FLJ42486	0.728	1.07E-06	0.00188
cg18236477	13	51761	ATP8A2	0.722	1.45E-06	0.00188
cg27320127	2	56660	KCNK12	0.722	1.44E-06	0.00188
cg07621046	10	143379	C10orf82	0.724	1.30E-06	0.00188
cg14456683	3	7545	ZIC1	0.709	2.71E-06	0.00277
cg21801378	15	60677	BRUNOL6	0.703	3.51E-06	0.00291
cg00107187	14	388021	FLJ42486	0.702	3.65E-06	0.00291
cg05508084	19	63934	ZNF667	0.704	3.37E-06	0.00291
cg07533148	1	25893	TRIM58	0.701	3.89E-06	0.00291
cg03975694	19	163255	ZNF540	0.700	4.06E-06	0.00291
cg19945840	1	126792	B3GALT6	0.699	4.30E-06	0.00293

FIG. 4B

cg00399483	18	1630	DCC	0.697	4.62E-06	0.00301
cg06291867	10	3363	HTR7	0.687	7.32E-06	0.00438
cg24826867	16	3394	IRF8	0.681	9.24E-06	0.00474
cg06092815	2	80309	SKIP	0.679	1.02E-05	0.00503
cg19885761	5	10814	CPLX2	0.677	1.10E-05	0.00526
cg24199834	4	5458	POU4F2	0.675	1.21E-05	0.00561
cg04528819	7	136259	KLF14	0.673	1.29E-05	0.00578
cg13434842	8	2626	GATA4	0.667	1.64E-05	0.00711
cg02008154	7	57057	TBX20	0.664	1.86E-05	0.00786
cg25044651	5	206338	FLJ90650	0.661	2.05E-05	0.00803
cg02994956	22	4744	NEFH	0.661	2.07E-05	0.00803
cg20366906	13	5100	PCDH8	0.656	2.50E-05	0.00944
cg27389185	19	163255	ZNF540	0.654	2.73E-05	0.01003
cg14826456	10	153	ADRB1	0.653	2.88E-05	0.01032
cg12111714	13	51761	ATP8A2	0.651	3.03E-05	0.01061
cg12457773	6	140767	VMP	0.649	3.36E-05	0.01147
cg11981599	8	2626	GATA4	0.646	3.67E-05	0.01166
cg00059225	5	2741	GLRA1	0.646	3.63E-05	0.01166

FIG. 4C

cg06572160	19	3748	KCNC3	0.642	4.21E-05	0.01262
cg08668790	19	7710	ZNF154	0.642	4.22E-05	0.01262
cg12782180	7	3952	LEP	0.641	4.43E-05	0.01296
cg20134215	6	84539	MCHR2	0.633	5.84E-05	0.01505
cg13921352	3	151647	FAM19A4	0.631	6.37E-05	0.01575
cg13614181	13	28984	RGC32	0.628	6.89E-05	0.01648
cg23563234	5	56099	PCDHGB7	0.628	7.06E-05	0.01661
cg00201234	3	2199	FBLN2	0.626	7.57E-05	0.01696
cg21992250	11	51296	SLC15A3	0.626	7.56E-05	0.01696
cg00911351	5	8641	PCDHGB4	0.625	7.73E-05	0.01705
cg19594666	7	3952	LEP	0.624	7.91E-05	0.01720
cg15425280	4	2891	GRIA2	0.619	9.39E-05	0.01898
cg06156376	3	6474	SHOX2	0.619	9.37E-05	0.01898
cg19831077	7	349136	LOC349136	0.618	9.77E-05	0.01946
cg27409364	11	3746	KCNC1	0.613	0.00011468	0.02137
cg15747595	8	85453	TSPYL5	0.612	0.000121753	0.02211
cg02154186	8	10687	PNMA2	0.608	0.000135948	0.02408
cg20616414	9	65268	WNK2	0.605	0.000148266	0.02502

FIG. 4D

cg10235817	4	152	ADRA2C	0.605	0.000151453	0.02526
cg16232126	2	60482	SLC5A7	0.602	0.00016384	0.02671
cg24646414	8	2626	GATA4	0.603	0.000162296	0.02671
cg17241310	1	343472	BARHL2	0.600	0.00017402	0.02750
cg10031651	3	79442	LRRC2	0.596	0.000199077	0.03104
cg23290344	8	4741	NEF3	0.595	0.000207878	0.03207
cg06908778	10	9576	SPAG6	0.591	0.000230493	0.03283
cg02844545	6	9247	GCM2	0.594	0.000215307	0.03283
cg14614211	10	283078	IRXL1	0.592	0.000224639	0.03283
cg25511429	6	51299	NRN1	0.589	0.000244086	0.03367
cg15201635	16	55512	SMPD3	0.590	0.000243284	0.03367
cg19246110	19	79891	ZNF671	0.589	0.000251007	0.03397
cg06760035	17	3214	HOXB4	0.587	0.000263627	0.03534
cg13603171	6	26002	MOXD1	0.583	0.000295067	0.03883
cg13282837	14	8115	TCL1A	0.579	0.00033117	0.04280
cg20792062	12	3741	KCNA5	0.571	0.000421691	0.04999
cg02228185	17	443	ASPA	-0.731	9.14E-07	0.00188
cg13547237	11	83638	Bles03	-0.717	1.84E-06	0.00218

FIG. 4E

cg09809672	1	128178	EDARADD	-0.715	1.98E-06	0.00218
cg01293143	20	6919	TCEA2	-0.703	3.61E-06	0.00291
cg05822532	7	2006	ELN	-0.688	7.00E-06	0.00437
cg07408456	19	114770	PGLYRP2	-0.685	7.90E-06	0.00453
cg08468689	17	84514	LGP1	-0.682	8.74E-06	0.00464
cg27210390	17	10040	TOM1L1	-0.683	8.53E-06	0.00464
cg01820374	12	3902	LAG3	-0.663	1.93E-05	0.00790
cg17589341	18	6563	SLC14A1	-0.640	4.63E-05	0.01302
cg19761273	17	1453	CSNK1D	-0.634	5.62E-05	0.01493
cg03440846	20	55902	ACSS2	-0.611	0.00012427	0.02228
cg08909157	9	157983	C9orf66	-0.607	0.00013896	0.02431
cg11136562	5	64411	CENTD3	-0.592	0.000227926	0.03283
cg16464322	19	3191	HNRPL	-0.592	0.000227715	0.03283
cg08872742	16	1003	CDH5	-0.582	0.00030785	0.04015
cg18328933	3	25864	ABHD14A	-0.579	0.000338098	0.04330
cg15784615	12	4055	LTBR	-0.577	0.000350304	0.04374
cg23282949	X	5973	REBP	-0.577	0.000358602	0.04435

TargetID	Chr	MapInfo	Gene_ID	Symbol	corr	qval
cg00059225	5	151284550	2741	GLRA1	0.646	0.012
cg12799895	7	98084588	4885	NPTX2	0.754	0.001
cg15747595	8	98359056	85453	TSPYL5	0.612	0.022
cg16232126	2	107969437	60482	SLC5A7	0.602	0.027
cg18236477	13	24941066	51761	ATP8A2	0.722	0.002
cg19594666	7	127668516	3952	LEP	0.624	0.017
cg19885761	5	175156252	10814	CPLX2	0.677	0.005
cg19945840	1	1157899	126792	B3GALT6	0.699	0.003
cg21801378	15	70399179	60677	BRUNOL6	0.703	0.003
cg27320127	2	47651900	56660	KCNK12	0.722	0.002

FIG. 5

Category	Significance	List of genes in each category
Cardiovascular disease	$p=1.59 \times 10^{-6}$	ADRA2C, ADRB1, ATP8A2, DCC, ELN, FAM19A4, GATA4, GRIA2, HTR7, IRF8, KCNA5, KCNC1, LEP, MOXD1, NRSN1, PDE4C, TBX20
Genetic disease	$p=1.59 \times 10^{-6}$	ADRA2C, ADRB1, ASPA, ATP8A2, CPLX2, EDARADD, ELN, GATA4, GCM2, GRIA2, HTR7, KCNA5, LEP, NEFH, NEFM, PCDH8, PDE4C, SLC14A1, SLC5A7, TBX20
Neurological disease	$p=1.47 \times 10^{-4}$	ADRA2C, ADRB1, ASPA, ATP8A2, CPLX2, GRIA2, HTR7, LEP, NEFH, NEFM, PCDH8, PDE4C, SLC14A1, SLC5A7
Molecular transport	$p=2.4 \times 10^{-3}$	ADRB1, CPLX2, GLRA1, HTR7, KCNA5, KCNC1, KCNC3, LEP, SLC14A1, SLC5A7

FIG. 6

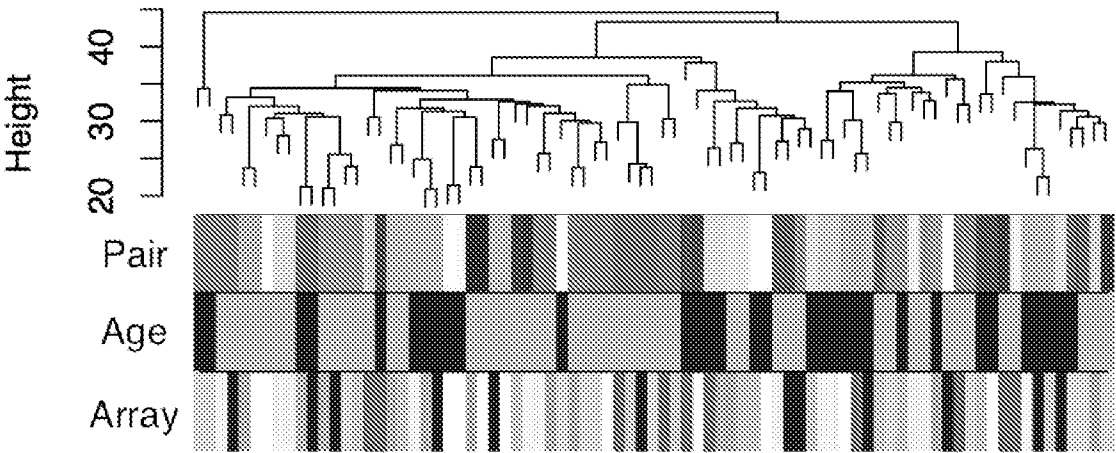


FIG. 7

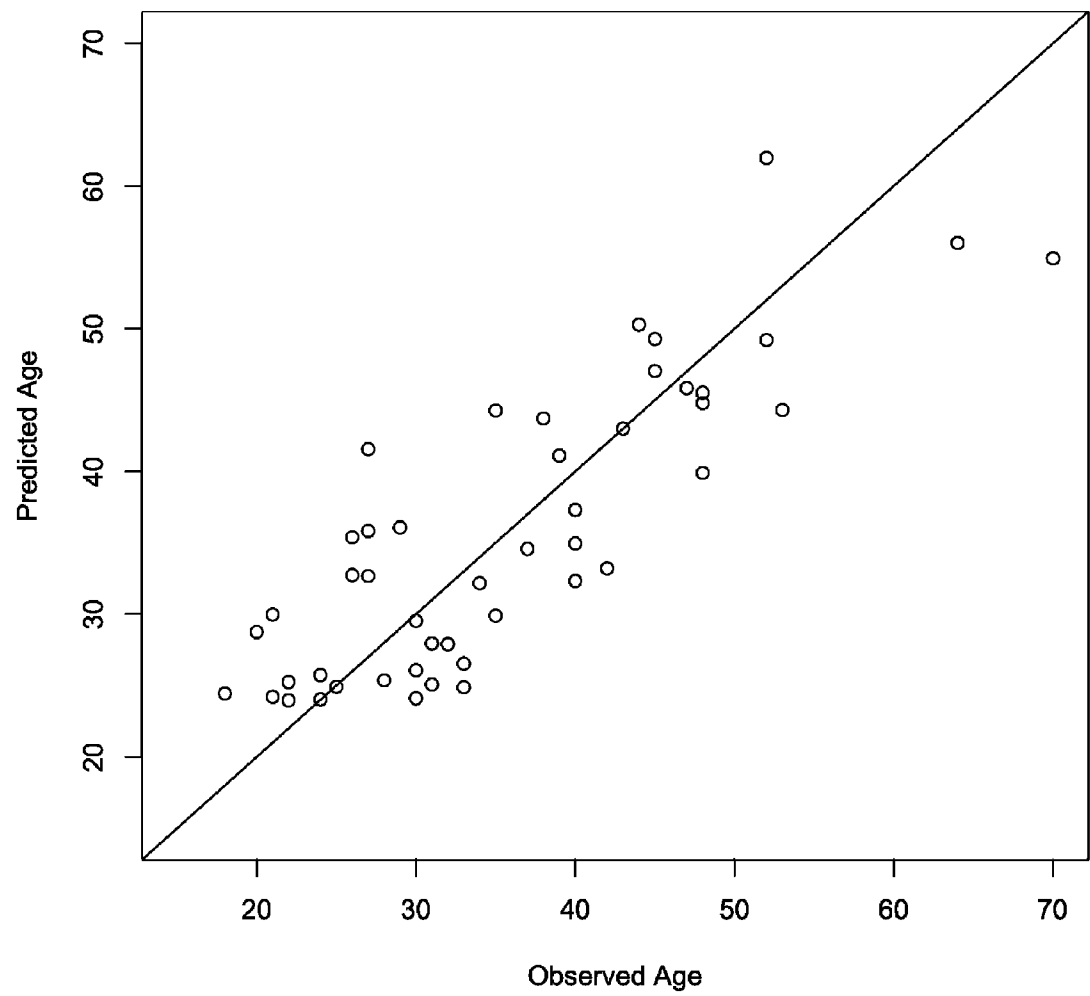


FIG. 8

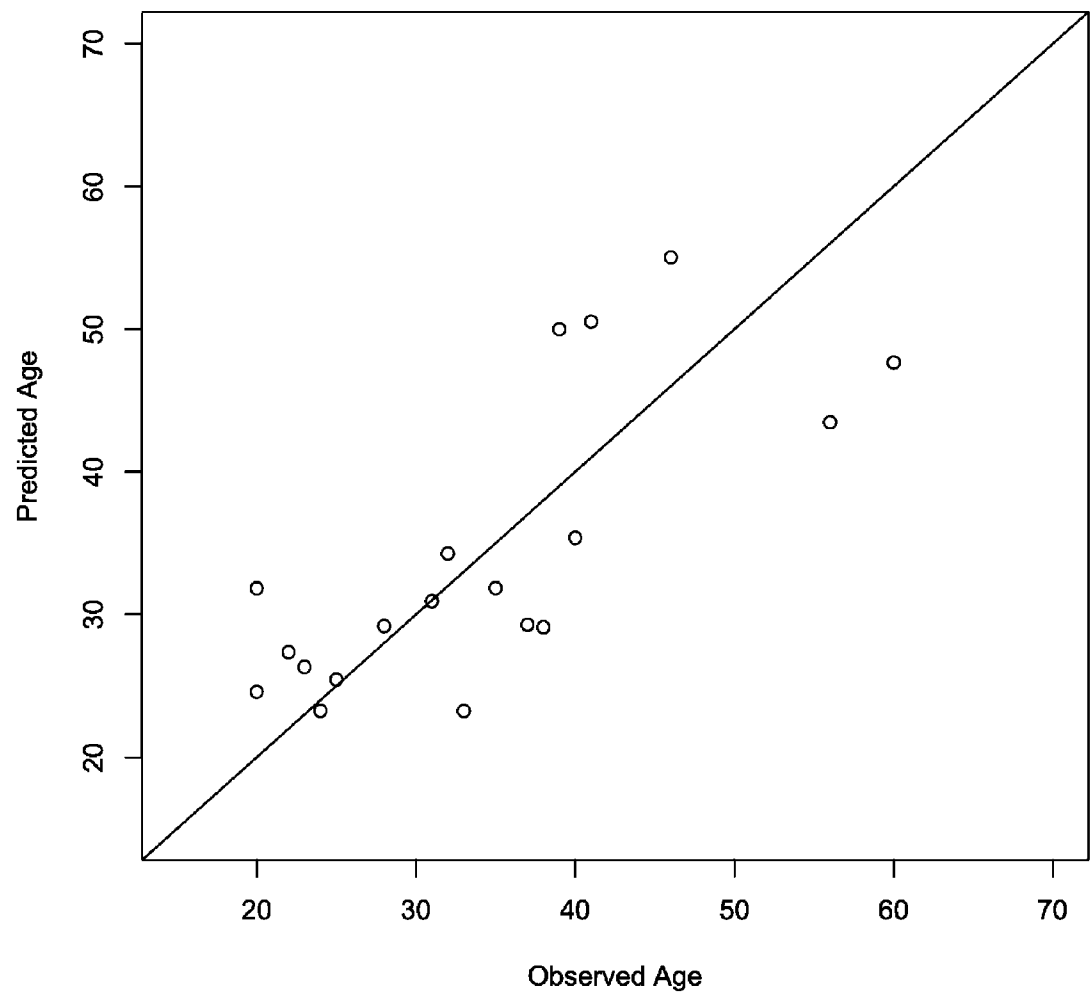


FIG. 9

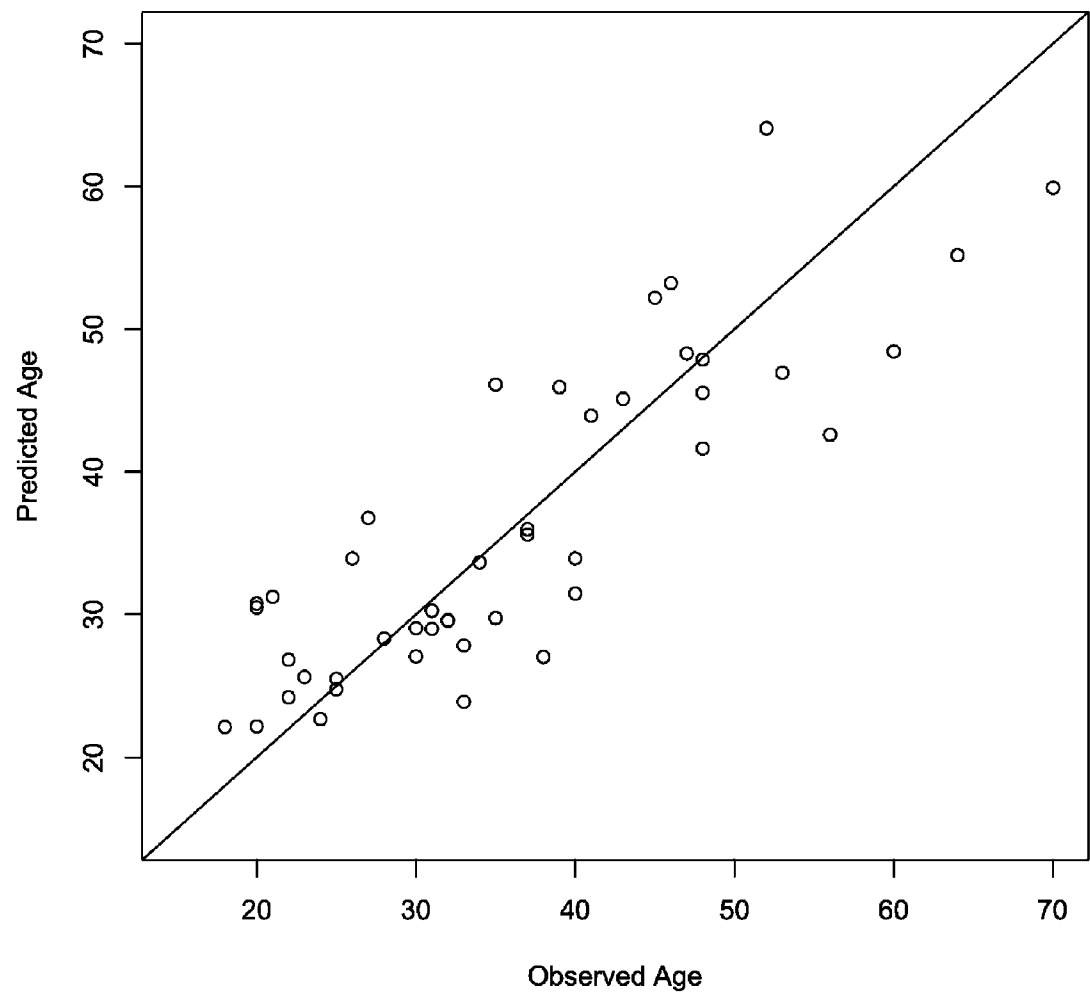


FIG. 10

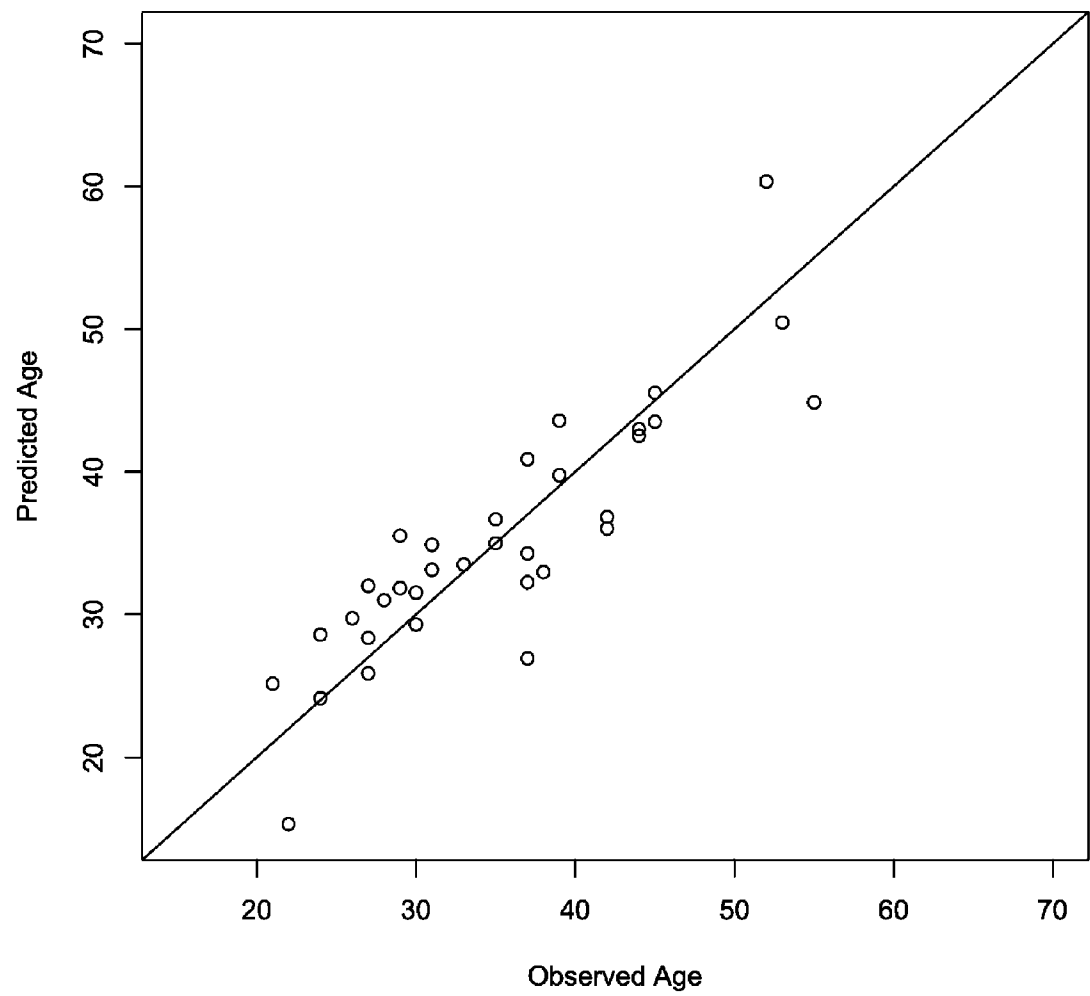


FIG. 11

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 2012/038552

A. CLASSIFICATION OF SUBJECT MATTER

C12Q 1/68 (2006.01)

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

C12Q 1/68

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

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

ARIPO, ATPO, AUPO, DPMA, DWPI, EBI, EPO, Esp@cenet GBPO, NCBI, OAPI, PAJ, USPTO, WIPO

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	JOAO PEDRO DE MAGALHAES et al.: "Meta-analysis of age-related gene expression profiles identifies common signatures of aging", Bioinformatics Advance Access published, 02.02.2009, p. 1-7, especially, p. 6, col. 1, paragraph 3	1-20
Y	US 2009/0263804 A1 (UNIVERSITY OF KENTUCKY RESEARCH FOUNDATION) 22.10.2009, paragraph [0107], table 1A, abstract	1-20
Y	DATABASE GenBank, GX282377, 12.12.2010	1-20
Y	US 2004/0132026 A1 (ALEXANDER OLEK) 08.07.2004, p. 2-4, claims	1-20
Y	WO 2003/025215 A1 (THE UNIVERSITY OF QUEENSLAND) 27.03.2003, p. 2, claims 40-69, 72-87	1-20
Y	DENA G. HERNANDEZ et al.: "Distinct DNA methylation changes highly correlated with chronological age in the human brain", Human Molecular Genetics, 2011, 20(6):1164-1172	1-20



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A" document defining the general state of the art which is not considered to be of particular relevance	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E" earlier document but published on or after the international filing date	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&" document member of the same patent family
"O" document referring to an oral disclosure, use, exhibition or other means	
"P" document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search

06 July 2012 (06.07.2012)

Date of mailing of the international search report

20 September 2012 (20.09.2012)

Name and mailing address of the ISA/ FIPS
Russia, 123995, Moscow, G-59, GSP-5,
Berezhkovskaya nab., 30-1

Facsimile No. +7 (499) 243-33-37

Authorized officer

M. Kuptsova

Telephone No. (495) 531-65-15

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 2012/038552

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

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

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

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

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

see extra sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.: **1-20 (partially)**

Remark on Protest

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

Continuation of Box III

The group of invention, which is represented in Claims a priory, does not meet the requirement of the invention unity in accordance with the Rule 13 since it is absent the peculiar common technical feature, uniting the declared inventions by means of single inventive concept.

The group of alternative features, which are represented in Claims 14 and in Claim 16 is not equivalent with respect to the group of the alternative features, which are represented in Claim 1, and is not a partial case of their embodiment, since a row of loci, which are listed in Claims 14 and 16, are not belonged to SEG ID NO: 1-4. As consequence, in a part of analysis only of CG-loci, which are not enter in the sequences SEG ID NO: 1-4, Claims 14 and 16 have no common peculiar peculiar features wit Claim 1.

Thus, a pattern of methylation of cytosine, at least, in one of sequences SEG ID NO: 1-4 is represented by means of peculiar technical feature, which unites by means of single inventive concept the object, which is declared in Claim 1, and the objects, that are declared in Claims 14 and 16, only a part of analysis, at least of CG-loci, which belong to the SEG ID NO: 1-4.

The pattern of methylation of cytosine in the loci, which are not belonged to the SEG ID NO: 1-4, unites a priory by means of single inventive concept the object only the objects, which are declared in Claims 14 and 16 in a part of analysis of the mentioned CG-loci.

Therefore, from the viewpoint of the Rule 13 in the represented Claims it may be separated the following two groups of the inventions, which are not united against each other by means of single inventive concept:

1 group - Claims 1 and Claims 14, 16 in a part of analysis, at least, of CG-loci, which are belonged to the SEG ID NO: 1-4;

2 group - Claims 14, 16 in a part of analysis, only of CG-loci, which are not belonged to the SEG ID NO: 1-4.