

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

(19) World Intellectual Property

Organization

International Bureau

(43) International Publication Date

24 August 2023 (24.08.2023)



(10) International Publication Number

WO 2023/158680 A1

(51) International Patent Classification:

G16B 20/00 (2019.01) C12Q 1/6827 (2018.01)

G16H 50/20 (2018.01) C12Q 1/6876 (2018.01)

G16H 50/30 (2018.01) C12Q 1/6883 (2018.01)

SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

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

(21) International Application Number:

PCT/US2023/013119

(22) International Filing Date:

15 February 2023 (15.02.2023)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

63/310,362 15 February 2022 (15.02.2022) US

(71) Applicant: THE BOARD OF TRUSTEES OF THE LE-

LAND STANFORD JUNIOR UNIVERSITY [US/US];
Office of the General Counsel, Building 170, 3rd
Floor, Main Quad, P.O. Box 20386, Stanford, California
94305-2038 (US).

(72) Inventors: **MOQRI, Mahdi**; 415 Broadway Street, 2nd
Floor, MC8854, Redwood City, California 94063 (US). **SE-**
BASTIANO, Vittorio; 1277 Laurel Street, Menlo Park,
California 94025 (US). **CIPRIANO, Andrea**; 3375 Alma
Street #367, Palo Alto, California 94306 (US). **SNYDER,**
Michael P.; 949 Cottrell Way, Stanford, California 94305
(US).

(74) Agent: **SHERWOOD, Pamela J.**; 201 Redwood Shores
Parkway, Suite 200, Redwood City, California 94065 (US).

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

(84) Designated States (unless otherwise indicated, for every
kind of regional protection available): ARIPO (BW, CV,
GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ,
TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU,
TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE,
DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU,
LV, MC, ME, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI,

(54) Title: AGEINDEX, A WHOLE-GENOME EPIGENETIC INDEX TO MEASURE MOLECULAR AGING AND REJUVE-
NATION

(57) Abstract: Compositions and methods are provided for determining a whole-genome epigenetic aging index, which epigenetic aging index is referred to herein as AgeIndex. AgeIndex quantifies biological aging at a cellular and chromosomal level, in different tissues, both in vitro and in vivo, and can be used to assess the efficacy of anti-aging and rejuvenation interventions.

WO 2023/158680 A1

AGEINDEX, A WHOLE-GENOME EPIGENETIC INDEX TO MEASURE MOLECULAR AGING AND REJUVENATION

CROSS REFERENCE TO OTHER APPLICATIONS

[0001] This application claims the benefit of U.S. Provisional Application No. 63/310,362, filed February 15, 2022, the contents of which are hereby incorporated by reference in its entirety.

BACKGROUND

[0002] DNA methylation (DNAm) is the epigenetic modification most strongly associated with aging. Numerous studies have identified examples of specific CpGs whose methylation levels are strongly correlated with age, which have been used to build epigenetic clocks. In parallel, several studies have demonstrated that there is a global increase in epigenetic entropy as a result of the loss of epigenetic information, which occurs when regions that are predominantly hypo- or hyper-methylated at birth gradually transition to a state of partial methylation. However, CpG-based clock methods suffer from limited mechanistic interpretability as the CpGs are chosen based purely on their correlation with age.

[0003] On the other hand, measures of epigenetic entropy assume that the methylation states trend towards partial methylation and increased randomness. This basic assumption does not hold for sites most highly correlated with age, as they transition from one methylation state toward the other.

[0004] The process of aging is a multifactorial phenomenon that involves still mostly unknown mechanisms, all of them having as outcome the loss of functionality of the affected organism. Measuring aging is essential in order to test the reliability and efficacy of any anti-aging/rejuvenation treatment. Despite the huge efforts in the field, how to measure biological aging still remains an open question. Epigenome is both actively and passively lost as we age and several epigenetic aging clocks have been developed recently. However, the current clocks are not tuned; for rejuvenation, therefore a robust measure to accurately quantify aging as well as rejuvenation has not been found yet.

SUMMARY

[0005] Compositions and methods are provided for determining a whole-genome epigenetic aging index based on whole genome analysis, which epigenetic aging index is referred to herein as AgeIndex. AgeIndex quantifies biological aging (also referred to herein as molecular aging) at a cellular and chromosomal level, in different tissues, both in vitro and in vivo, and can be used to assess the efficacy of anti-aging and rejuvenation treatments.

[0006] AgeIndex provides an unbiased method to estimate the age of a cell by measuring the loss of its epigenetic information. As cells replicate, the transfer of epigenetic information to

cells is not completely faithful, therefore, cellular replication accompanies a gradual loss of important information. AgeIndex captures bidirectional transitions in genome-wide methylation states. These transitions reflect methylation gain in unmethylated, CpG-dense regions enriched for EZH2 or EHMT2 binding, and methylation loss in methylated, CpG-sparse regions associated with lamin. The age-dependent loss of epigenetic information can be measured by comparing a cell with a control cell or reference, e.g. an embryonic stem cell (defined as Age 0 cells). AgeIndex can also be applied to free circulating DNA to provide a fast and reliable method to assess biological aging.

[0007] In some embodiments, the features that are analyzed for the presence of methylation are: lamin-associated domains (LADs); CpG-protected regions, defined as CpG density divided by TpG+CpA density; regions with EZH2 DNA-binding domains; and regions with EHMT2 DNA-binding domains. Features may also comprise regions that are highly methylated at birth; and regions that are low methylated at birth.

[0008] It is observed that over aging, there is a global loss of DNA methylation at lamin-associated regions; a global gain of DNA methylation in CpG-rich regions, defined as CpG-protected regions; and EZH2 and EHMT2 DNA-binding domains. High methylated regions at birth lose methylation across the lifespan. Conversely, DNA methylation levels of low methylated regions at birth increase with age.

[0009] These features identified above as associated with the age-dependent changes in the epigenome are combined to capture the epigenetic state of cells in a holistic approach, and are integrated in a machine learning algorithm that measures the epigenetic loss of information by capturing genome-wide gain or loss of DNA methylation occurring during age. Genomic features are divided into two groups based on their association with the gain or loss of DNA methylation, where age-associated loss of methylation is found with (a) high methylated regions at birth; and (b) lamin-associated regions. Age-associated gain of methylation is observed with (a) CpG-protected regions; (b) EZH2 DNA-binding domains; (c) EHMT2 DNA-binding domains; and (d) low-methylated regions at birth. The DNA methylation levels are aggregated across the features in the two groups, providing AgeIndex_gain and AgeIndex_loss indices. The large numbers of CpG sites in each group makes AgeIndex an assay-agnostic method, and allows aggregation of changes in methylation across different types of genomic regions.

[0010] In some embodiments AgeIndex is used to perform a metagene analysis of loss and gain of epigenetic information around the transcription start site (TSS) of genes. In some embodiments AgeIndex is used to perform a metagene analysis of loss and gain of epigenetic information of individual chromosomes. In some embodiments AgeIndex is used to perform a metagene analysis of loss and gain of epigenetic information in specific tissues. In some embodiments AgeIndex is used to perform a metagene analysis of loss and gain of epigenetic

information in cancer cells. In some embodiments AgeIndex is used to perform a metagene analysis of loss and gain of epigenetic information to track changes of aging intervention treatments.

[0011] In other embodiments, the method is executed through the use of a computer based software program wherein the methylation features are inputted and the software program outputs a score indicative of a particular classification as defined by the user. The software programs employs machine learning to uncover relationships between input metrics in their relation to target outputs through training algorithms.

[0012] In one embodiment of the invention, the methods of determining an analysis of molecular aging in an individual, comprising obtaining a patient sample comprising patient DNA, e.g. cells or circulating DNA. Skin samples and blood samples are a convenient source of cells and DNA.

[0013] Also described herein is a method for generating an AgeIndex, comprising: obtaining a dataset associated with a sample obtained from the subject, wherein the dataset comprises quantitative data from the methylation markers disclosed herein and analyzing the dataset classification relative to the AgeIndex model, wherein a statistically significant match with a model disclosed herein is indicative of the molecular age of the sample. The data may be analyzed by a computer processor. The processor may be communicatively coupled to a storage memory for analyzing the data. The processor may be coupled to a sequencer, and may include algorithms for predictive classification. Also described herein is a computer-readable storage medium storing computer-executable program code, the program code comprising: program code for storing and analyzing data obtained by the methods of the disclosure.

[0014] In other embodiments of the invention a device or kit is provided for the analysis of patient samples. Such devices or kits will include reagents that specifically identify one or more cells, indicative of the status of the patient, including without limitation affinity reagents. The reagents can be provided in isolated form, or pre-mixed as a cocktail suitable for the methods of the invention. A kit can include instructions for using the plurality of reagents to determine data from the sample; and instructions for statistically analyzing the data. The kits may be provided in combination with a system for analysis, e.g. a system implemented on a computer. Such a system may include a software component configured for analysis of data obtained by the methods of the invention. In some embodiments, machine learning tools for epigenetic analytics are used to quantify how a DNA population varies in biological aging. The features quantified in these analyses can be used to build a random forest classifier for predicting biological age.

BRIEF DESCRIPTION OF THE DRAWINGS

[0015] The invention is best understood from the following detailed description when read in conjunction with the accompanying drawings. It is emphasized that, according to common practice, the various features of the drawings are not to-scale. On the contrary, the dimensions of the various features are arbitrarily expanded or reduced for clarity. Included in the drawings are the following figures.

[0016] **FIGS. 1A-1E.** Analyses based on all CpGs covered in Methyl450, excluding SNPs and sex chromosomes. **a**, Correlation between DNA methylation levels (in neonatal blood samples, GSE103657) and the genomic distribution of each dinucleotide (within 1 Kb around the focal CpG). Different dinucleotide-densities associate with low or high DNA methylation levels. DNA methylation levels are lowest at CpG-rich regions and highest at TG/CA-rich regions. **b**, Comparison of CpG Island and “CpG-protection.” CpG-protection is more strongly associated with low DNA methylation. **c**, TF-binding domains, from ENCODE3, with the highest methylation levels within CGIs (in neonatal blood samples. EZH2 and EHMT2 DNA-binding sites are among the most highly methylated regions within CGIs for consistency). **d**, Comparison of DNA methylation levels in neonates and adults (GSE152027). Globally, CpG-rich regions gain DNA methylation and lamin-associated CpG-poor regions (Zhou et al. 2018) lose DNA methylation. **e**, The effect of each feature on age-dependent changes in DNA methylation, controlling for other features. Each genomic feature (e.g. CpG-density and EZH2/EHMT2 DNA-binding) independently contribute to the gain of DNA methylation.

[0017] **FIGS. 2A-2D.** **a**, DNA methylation levels at late-replicating regions in blood samples across different ages. Blood samples from four different Methyl450 datasets, spanning from 13 to 100 years old. Data confirms a global loss of DNA methylation in lamin-associated CpG-poor regions (Zhou et al. 2018) **b**, The initial state of DNA methylation at birth associate with the direction of age-dependent change in methylation, in blood. Lowly methylated regions gain DNA methylation while highly methylated regions lose DNA methylation. **c**, CG-rich regions, gain DNA methylation by age while TG/CA-rich regions lose DNA methylation. **d**, EZH2 and EHMT2 DNA binding regions consistently gain DNA methylation by age.

[0018] **FIGS. 3A-3B.** **a**, AgeIndex differentiates blood samples prepared using different assays, Methyl450 (GSE40279), EPIC (GSE132203), and WGBS (CD4 from GSE31263 and CD4 from BLUEPRINT). Overall, average DNAm decreases in AgeIndex_loss regions and increases in AgeIndex_gain regions, with aging. At the population level, AgeIndex can differentiate age groups with 10-year increments in samples prepared with Methyl450 and 5 years in samples prepared with EPIC. In addition, both AgeIndex_gain and loss can differentiate blood samples from children, middle-aged, and old adults, assayed using WGBS, at the chromosome level. **b**, AgeIndex differentiates sorted blood cells (neutrophils, monocytes, macrophages, and B cells) from different age groups (BLUEPRINT). For

Methyl450 analyses, GSExxx is used as the reference to identify methylation states at birth and for EPIC and WGBS analyses, GSExxx is used as the reference.

- [0019] FIG. 4. Change in AgeIndex with age.
- [0020] FIG. 5A-5D. Changes in AgeIndex across tissue types.
- [0021] FIG. 6. Changes in AgeIndex in epidermal cells.
- [0022] FIG. 7. Changes in AgeIndex for specific chromosomes of epidermal cells.
- [0023] FIG. 8. AgeIndex measurement of aging intervention.
- [0024] FIG. 9. AgeIndex analysis of normal and cancer cells.

DETAILED DESCRIPTION

- [0025] Before the present methods and compositions are described, it is to be understood that this invention is not limited to particular method or composition described, as such may, of course, vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to be limiting, since the scope of the present invention will be limited only by the appended claims.
- [0026] Where a range of values is provided, it is understood that each intervening value, to the tenth of the unit of the lower limit unless the context clearly dictates otherwise, between the upper and lower limits of that range is also specifically disclosed. Each smaller range between any stated value or intervening value in a stated range and any other stated or intervening value in that stated range is encompassed within the invention. The upper and lower limits of these smaller ranges may independently be included or excluded in the range, and each range where either, neither or both limits are included in the smaller ranges is also encompassed within the invention, subject to any specifically excluded limit in the stated range. Where the stated range includes one or both of the limits, ranges excluding either or both of those included limits are also included in the invention.
- [0027] Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, some potential and preferred methods and materials are now described. All publications mentioned herein are incorporated herein by reference to disclose and describe the methods and/or materials in connection with which the publications are cited. It is understood that the present disclosure supercedes any disclosure of an incorporated publication to the extent there is a contradiction.
- [0028] It must be noted that as used herein and in the appended claims, the singular forms "a", "an", and "the" include plural referents unless the context clearly dictates otherwise. Thus, for example, reference to "a cell" includes a plurality of such cells and reference to "the

peptide" includes reference to one or more peptides and equivalents thereof, e.g. polypeptides, known to those skilled in the art, and so forth.

[0029] The publications discussed herein are provided solely for their disclosure prior to the filing date of the present application. Nothing herein is to be construed as an admission that the present invention is not entitled to antedate such publication by virtue of prior invention. Further, the dates of publication provided may be different from the actual publication dates which may need to be independently confirmed.

[0030] As used herein, compounds which are "commercially available" may be obtained from commercial sources including but not limited to Acros Organics (Pittsburgh PA), Aldrich Chemical (Milwaukee WI, including Sigma Chemical and Fluka), Apin Chemicals Ltd. (Milton Park UK), Avocado Research (Lancashire U.K.), BDH Inc. (Toronto, Canada), Bionet (Cornwall, U.K.), Chemservice Inc. (West Chester PA), Crescent Chemical Co. (Hauppauge NY), Eastman Organic Chemicals, Eastman Kodak Company (Rochester NY), Fisher Scientific Co. (Pittsburgh PA), Fisons Chemicals (Leicestershire UK), Frontier Scientific (Logan UT), ICN Biomedicals, Inc. (Costa Mesa CA), Key Organics (Cornwall U.K.), Lancaster Synthesis (Windham NH), Maybridge Chemical Co. Ltd. (Cornwall U.K.), Parish Chemical Co. (Orem UT), Pfaltz & Bauer, Inc. (Waterbury CN), Polyorganix (Houston TX), Pierce Chemical Co. (Rockford IL), Riedel de Haen AG (Hannover, Germany), Spectrum Quality Product, Inc. (New Brunswick, NJ), TCI America (Portland OR), Trans World Chemicals, Inc. (Rockville MD), Wako Chemicals USA, Inc. (Richmond VA), Novabiochem and Argonaut Technology.

[0031] Compounds can also be made by methods known to one of ordinary skill in the art. As used herein, "methods known to one of ordinary skill in the art" may be identified through various reference books and databases. Suitable reference books and treatises that detail the synthesis of reactants useful in the preparation of compounds of the present invention, or provide references to articles that describe the preparation, include for example, "Synthetic Organic Chemistry", John Wiley & Sons, Inc., New York; S. R. Sandler et al., "Organic Functional Group Preparations," 2nd Ed., Academic Press, New York, 1983; H. O. House, "Modern Synthetic Reactions", 2nd Ed., W. A. Benjamin, Inc. Menlo Park, Calif. 1972; T. L. Gilchrist, "Heterocyclic Chemistry", 2nd Ed., John Wiley & Sons, New York, 1992; J. March, "Advanced Organic Chemistry: Reactions, Mechanisms and Structure", 4th Ed., Wiley-Interscience, New York, 1992. Specific and analogous reactants may also be identified through the indices of known chemicals prepared by the Chemical Abstract Service of the American Chemical Society, which are available in most public and university libraries, as well as through on-line databases (the American Chemical Society, Washington, D.C., www.acs.org may be contacted for more details). Chemicals that are known but not commercially available in catalogs may be prepared by custom chemical synthesis houses,

where many of the standard chemical supply houses (e.g., those listed above) provide custom synthesis services.

[0032] The terms "polypeptide," "peptide" and "protein" are used interchangeably herein to refer to a polymer of amino acid residues. The terms also apply to amino acid polymers in which one or more amino acid residue is an artificial chemical mimetic of a corresponding naturally occurring amino acid, as well as to naturally occurring amino acid polymers and non-naturally occurring amino acid polymer.

[0033] The term "sequence identity," as used herein in reference to polypeptide or DNA sequences, refers to the subunit sequence identity between two molecules. When a subunit position in both of the molecules is occupied by the same monomeric subunit (e.g., the same amino acid residue or nucleotide), then the molecules are identical at that position. The similarity between two amino acid or two nucleotide sequences is a direct function of the number of identical positions. In general, the sequences are aligned so that the highest order match is obtained. If necessary, identity can be calculated using published techniques and widely available computer programs, such as the GCS program package (Devereux et al., *Nucleic Acids Res.* 12:387, 1984), BLASTP, BLASTN, FASTA (Atschul et al., *J. Molecular Biol.* 215:403, 1990).

[0034] By "protein variant" or "variant protein" or "variant polypeptide" herein is meant a protein that differs from a wild-type protein by virtue of at least one amino acid modification. The parent polypeptide may be a naturally occurring or wild-type (WT) polypeptide, or may be a modified version of a WT polypeptide. Variant polypeptide may refer to the polypeptide itself, a composition comprising the polypeptide, or the amino sequence that encodes it. Preferably, the variant polypeptide has at least one amino acid modification compared to the parent polypeptide, e.g. from about one to about ten amino acid modifications, and preferably from about one to about five amino acid modifications compared to the parent.

[0035] By "parent polypeptide", "parent protein", "precursor polypeptide", or "precursor protein" as used herein is meant an unmodified polypeptide that is subsequently modified to generate a variant. A parent polypeptide may be a wild-type (or native) polypeptide, or a variant or engineered version of a wild-type polypeptide. Parent polypeptide may refer to the polypeptide itself, compositions that comprise the parent polypeptide, or the amino acid sequence that encodes it.

[0036] The term "amino acid" refers to naturally occurring and synthetic amino acids, as well as amino acid analogs and amino acid mimetics that function in a manner similar to the naturally occurring amino acids. Naturally occurring amino acids are those encoded by the genetic code, as well as those amino acids that are later modified, e.g., hydroxyproline, gamma-carboxyglutamate, and O-phosphoserine. "Amino acid analogs" refers to compounds that have the same basic chemical structure as a naturally occurring amino acid, i.e., an α -

carbon that is bound to a hydrogen, a carboxyl group, an amino group, and an R group, e.g., homoserine, norleucine, methionine sulfoxide, methionine methyl sulfonium. Such analogs have modified R groups (e.g., norleucine) or modified peptide backbones, but retain the same basic chemical structure as a naturally occurring amino acid. "Amino acid mimetics" refers to chemical compounds that have a structure that is different from the general chemical structure of an amino acid, but that functions in a manner similar to a naturally occurring amino acid.

[0037] Amino acid modifications disclosed herein may include amino acid substitutions, deletions and insertions, particularly amino acid substitutions. Variant proteins may also include conservative modifications and substitutions at other positions of the cytokine and/or receptor (e.g., positions other than those involved in the affinity engineering). Such conservative substitutions include those described by Dayhoff in *The Atlas of Protein Sequence and Structure* 5 (1978), and by Argos in *EMBO J.*, 8:779-785 (1989). For example, amino acids belonging to one of the following groups represent conservative changes: Group I: Ala, Pro, Gly, Gln, Asn, Ser, Thr; Group II: Cys, Ser, Tyr, Thr; Group III: Val, Ile, Leu, Met, Ala, Phe; Group IV: Lys, Arg, His; Group V: Phe, Tyr, Trp, His; and Group VI: Asp, Glu. Further, amino acid substitutions with a designated amino acid may be replaced with a conservative change.

[0038] The term "isolated" refers to a molecule that is substantially free of its natural environment. For instance, an isolated protein is substantially free of cellular material or other proteins from the cell or tissue source from which it is derived. The term refers to preparations where the isolated protein is sufficiently pure to be administered as a therapeutic composition, or at least 70% to 80% (w/w) pure, more preferably, at least 80%-90% (w/w) pure, even more preferably, 90-95% pure; and, most preferably, at least 95%, 96%, 97%, 98%, 99%, or 100% (w/w) pure. A "separated" compound refers to a compound that is removed from at least 90% of at least one component of a sample from which the compound was obtained. Any compound described herein can be provided as an isolated or separated compound.

[0039] The terms "subject," "individual," and "patient" are used interchangeably herein to refer to a mammal being assessed for treatment and/or being treated. In some embodiments, the mammal is a human. The terms "subject," "individual," and "patient" encompass, without limitation, individuals having a disease. Subjects may be human, but also include other mammals, particularly those mammals useful as laboratory models for human disease, e.g., mice, rats, etc.

[0040] The term "sample" with reference to a patient encompasses blood and other liquid samples of biological origin, solid tissue samples such as a biopsy specimen or tissue cultures or cells derived therefrom and the progeny thereof. The term also encompasses samples that have been manipulated in any way after their procurement, such as by treatment with

reagents; washed; or enrichment for certain cell populations, such as diseased cells. The definition also includes samples that have been enriched for particular types of molecules, e.g., nucleic acids, polypeptides, etc. The term "biological sample" encompasses a clinical sample, and also includes tissue obtained by surgical resection, tissue obtained by biopsy, cells in culture, cell supernatants, cell lysates, tissue samples, organs, bone marrow, blood, plasma, serum, and the like. A "biological sample" includes a sample obtained from a patient's diseased cell, e.g., a sample comprising polynucleotides and/or polypeptides that is obtained from a patient's diseased cell (e.g., a cell lysate or other cell extract comprising polynucleotides and/or polypeptides); and a sample comprising diseased cells from a patient. A biological sample comprising a diseased cell from a patient can also include non-diseased cells.

[0041] The term "diagnosis" is used herein to refer to the identification of a molecular or pathological state, disease or condition in a subject, individual, or patient.

[0042] The term "prognosis" is used herein to refer to the prediction of the likelihood of an event such as age-associated conditions, including recurrence, spread, and drug resistance, in a subject, individual, or patient. The term "prediction" is used herein to refer to the act of foretelling or estimating, based on observation, experience, or scientific reasoning, the likelihood of a subject, individual, or patient experiencing a particular event or clinical outcome. In one example, a physician may attempt to predict the likelihood that a patient will survive.

[0043] As used herein, the terms "treatment," "treating," and the like, refer to administering an agent, or carrying out a procedure, for the purposes of obtaining an effect on or in a subject, individual, or patient. The effect may be prophylactic in terms of completely or partially preventing a disease or symptom thereof and/or may be therapeutic in terms of effecting a partial or complete cure for a disease and/or symptoms of the disease. "Treatment," as used herein, may include treatment of cancer in a mammal, particularly in a human, and includes: (a) inhibiting the disease, i.e., arresting its development; and (b) relieving the disease or its symptoms, i.e., causing regression of the disease or its symptoms.

[0044] "In combination with", "combination therapy" and "combination products" refer, in certain embodiments, to the concurrent administration to a patient of the engineered proteins and cells described herein in combination with additional therapies, e.g. surgery, radiation, chemotherapy, and the like. When administered in combination, each component can be administered at the same time or sequentially in any order at different points in time. Thus, each component can be administered separately but sufficiently closely in time so as to provide the desired therapeutic effect.

[0045] "Concomitant administration" means administration of one or more components, such as engineered proteins and cells, known therapeutic agents, etc. at such time that the combination will have a therapeutic effect. Such concomitant administration may involve

concurrent (*i.e.* at the same time), prior, or subsequent administration of components. A person of ordinary skill in the art would have no difficulty determining the appropriate timing, sequence and dosages of administration.

[0046] The use of the term "in combination" does not restrict the order in which prophylactic and/or therapeutic agents are administered to a subject with a disorder. A first prophylactic or therapeutic agent can be administered prior to (e.g., 5 minutes, 15 minutes, 30 minutes, 45 minutes, 1 hour, 2 hours, 4 hours, 6 hours, 12 hours, 24 hours, 48 hours, 72 hours, 96 hours, 1 week, 2 weeks, 3 weeks, 4 weeks, 5 weeks, 6 weeks, 8 weeks, or 12 weeks before), concomitantly with, or subsequent to (e.g., 5 minutes, 15 minutes, 30 minutes, 45 minutes, 1 hour, 2 hours, 4 hours, 6 hours, 12 hours, 24 hours, 48 hours, 72 hours, 96 hours, 1 week, 2 weeks, 3 weeks, 4 weeks, 5 weeks, 6 weeks, 8 weeks, or 12 weeks after) the administration of a second prophylactic or therapeutic agent to a subject with a disorder.

[0047] *Aging* can be defined as chronologic time; or can be defined according to molecular changes associated with chronologic aging. Biologic (molecular) aging may refer to the time-related deterioration of the physiological functions necessary for survival and fertility. At the biological level, aging results from the impact of the accumulation of a wide variety of molecular and cellular damage over time. This leads to a gradual decrease in physical and mental capacity, a growing risk of disease and ultimately death. These changes are neither linear nor consistent, and they are only loosely associated with a person's age in years. The diversity seen in older age is not random.

[0048] In some biological pathways, functional decline can be defined in a mono-causal way, such as the decline of resting metabolism, whereas in other pathways the scope of the decline is rather broad and elusive, such as that for reduced stability of epigenetic patterns. Although epigenetic patterns change dramatically during development, these early events are biologically programmed and necessary, whereas alterations of the epigenome in adult somatic tissue may reflect aging-associated deleterious events.

[0049] The maximum life span is a characteristic of the species. It is the maximum number of years a member of that species has been known to survive. The maximum human life span is estimated to be 121 years. The life expectancy, the amount of time a member of a species can expect to live, is not characteristic of species, but of populations. It is usually defined as the age at which half the population still survives.

[0050] *DNA methylation.* There is a relationship between aging and changes in DNA methylation. DNA methylation patterns are shaped by two opposing processes of adding and removing a methyl group at position five of cytosine in DNA. Selective maintenance of DNA methylation at specific loci is essential for controlling differential expression of the paternal

and maternal alleles in mammals, known as genomic imprinting. After the developmental phase, the genome of somatic cells will consist of roughly 1% methylated DNA cytosines. While there exists great variability between the established patterns of DNA methylation, there is a consistent landmark in the form of CpG islands, which are unmethylated GC-rich regions with high densities of CpGs and often correlated with promoter regions.

[0051] The enzymes that transfer a methyl group from SAM to DNA producing 5-methylcytosine, are the family of DNA methyltransferases (DNMTs) that include DNMT1, DNMT3A, DNMT3B and DNMT3L. DNMT1 plays an important role in maintaining genomic methylation patterns. The activity of DNMT1 seems to decrease significantly during aging, which may be related to the global decrease of DNA methylation observed during aging. Another potential mechanism for DNA demethylation during aging is by enzymatic DNA demethylation catalyzed by the 5mC dioxygenases Ten-eleven translocation 1, 2, and 3 (TET1/2/3).

[0052] Methods to identify and quantify DNA methylation include: sodium bisulfite conversion and sequencing, differential enzymatic cleavage of DNA, and affinity capture of methylated DNA. Affinity-capture and bisulphite conversion followed by sequencing methods are generally used for gene specific or genome-wide analysis. The most commonly reported DNA affinity capture method is methylated DNA immunoprecipitation (Me-DIP) that uses methyl DNA specific antibody, or methyl capture using methyl-CpG binding domain (MBD) proteins. See, for example, Laird PW. (2010) *Nat. Rev. Genet.* 11:191-203; Beck, Stephan. (2010) *Nature Biotechnology*, 28, 1026-1028; and Nair, Shalima S. et al. (2011) *Epigenetics*, 6:1, 34-44; each specifically incorporated by reference.

[0053] Bisulfite genomic sequencing is regarded as a gold-standard technology for detection of DNA methylation because it provides a qualitative, quantitative and efficient approach to identify 5-methylcytosine at single base-pair resolution. It is based on the finding that the amination reactions of cytosine and 5-methylcytosine (5mC) proceed with very different consequences after the treatment of sodium bisulfite. In this regard, cytosines in single-stranded DNA are converted into uracil residues and recognized as thymine in subsequent PCR amplification and sequencing, however, 5mCs are immune to this conversion and remain as cytosines, allowing 5mCs to be distinguished from unmethylated cytosines. The actual methylation status can be determined either through direct PCR product sequencing (detection of average methylation status) or sub-cloning sequencing (detection of single molecules distribution of methylation patterns).

[0054] The methods of the present disclosure involve sequencing target loci following sodium bisulfite conversion, as well as analyzing sequence data. Various methods and protocols for DNA sequencing and analysis are well-known in the art and are described herein. For example, DNA sequencing may be accomplished using high-throughput DNA sequencing

techniques. Examples of next generation and high-throughput sequencing include, for example, massively parallel signature sequencing, polony sequencing, 454 pyrosequencing, Illumina (Solexa) sequencing with HiSeq, MiSeq, and other platforms, SOLiD sequencing, ion semiconductor sequencing (Ion Torrent), DNA nanoball sequencing, heliscope single molecule sequencing, single molecule real time (SMRT) sequencing, MassARRAY®, and Digital Analysis of Selected Regions (DANSR™). See, e.g., Stein RA (1 September 2008). "Next-Generation Sequencing Update". Genetic Engineering & Biotechnology News 28 (15); Quail, Michael; Smith, Miriam E; Coupland, Paul; Otto, Thomas D; Harris, Simon R; Connor, Thomas R; Bertoni, Anna; Swerdlow, Harold P; Gu, Yong (1 January 2012). "A tale of three next generation sequencing platforms: comparison of Ion torrent, pacific biosciences and illumina MiSeq sequencers". BMC Genomics 13 (1): 341 ; Liu, Lin; Li, Yinhu; Li, Siliang; Hu, Ni; He, Yimin; Pong, Ray; Lin, Danni; Lu, Lihua; Law, Maggie (1 January 2012). "Comparison of Next-Generation Sequencing Systems". Journal of Biomedicine and Biotechnology 2012: 1-11; Qualitative and quantitative genotyping using single base primer extension coupled with matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MassARRAY®). Methods Mol Biol. 2009;578:307-43; Chu T, Bunce K, Hogge WA, Peters DG. A novel approach toward the challenge of accurately quantifying fetal DNA in maternal plasma. Prenat Diagn 2010;30: 1226-9; and Suzuki N, Kamataki A, Yamaki J, Homma Y. Characterization of circulating DNA in healthy human plasma. Clinica chimica acta; international journal of clinical chemistry 2008;387:55-8). Similarly, software programs for primary and secondary analysis of sequence data are well-known in the art.

[0055] In some embodiments, high throughput sequencing generates at least 1,000, at least 5,000, at least 10,000, at least 20,000, at least 30,000, at least 40,000, at least 50,000, at least 100,000 or at least 500,000 sequence reads per hour; with each read being at least 50, at least 60, at least 70, at least 80, at least 90, at least 100, at least 120 or at least 150 bases per read. Sequencing can be performed using nucleic acids described herein such as genomic DNA, cDNA derived from RNA transcripts or RNA as a template. Sequencing may comprise massively parallel sequencing.

[0056] The methods disclosed herein may comprise amplification of DNA. Amplification may comprise PCR-based amplification. Alternatively, amplification may comprise nonPCR-based amplification. Amplification of DNA may comprise using bead amplification followed by fiber optics detection as described in Marguiles et al. "Genome sequencing in microfabricated high-density picolitre reactors", Nature, doi: 10.1038/nature03959; and well as in US Publication Application Nos. 200200 12930; 20030058629; 20030 1001 02; 20030 148344 ; 20040248 161 ; 200500795 10, 20050 124022; and 20060078909. Amplification of the nucleic acid may comprise use of one or more polymerases. The polymerase may be a DNA polymerase. The polymerase may be a RNA polymerase. The polymerase may be a high fidelity polymerase.

The polymerase may be KAPA HiFi DNA polymerase. The polymerase may be Phusion DNA polymerase. Amplification may comprise 20 or fewer amplification cycles. Amplification may comprise 20, 19, 18, 17, 16, 15, 14, 13, 12, 11, 10, or 9 or fewer amplification cycles. Amplification may comprise 18 or fewer amplification cycles. Amplification may comprise 16 or fewer amplification cycles. Amplification may comprise 15 or fewer amplification cycles.

[0057] Genomic features of interest include CpG-protected regions, which are defined as CpG density divided by TpG+CpA density. CpG islands may be defined as regions with: a length greater than 200bp, usually greater than about 500 bp, excluding repetitive Alu-elements, with a G+C content greater than 50%, and a ratio of observed to expected CpG greater than 0.6. Excluding repeated sequences, there are around 25,000 CpG islands in the human genome, 75% of which being less than 850bp long. There are greater than about 28,890 CpG islands in the human genome.

[0058] Regions with EHMT2 DNA-binding domains. Histone-lysine N-methyltransferase EHMT2 mono- and dimethylates "Lys-9" of histone H3 (H3K9me1 and H3K9me2, respectively) in euchromatin. H3K9me is a specific tag for repression of epigenetic transcription, recruiting HP1 proteins to methylated histones. EHMT2 also monomethylates "Lys-56" of histone H3 (H3K56me1) during the G1 phase of the cell cycle, and methylates "Lys-27" of histone H3 (H3K27me). EHMT2 also dimethylates other proteins, including "Lys-373" of p53. EHMT2 forms a heterodimer with EHMT1, and is a component of the E2F6.com-1 complex in G0 phase. EHMT2 contains seven ankyrin repeats which bind the monomethylated RELA subunit of NF-kappa-B, a SET domain which interacts with WIZ, and a pre-SET domain that binds three zinc ions via cysteine residues. Regions of EHMT2 DNA-binding domains may be determined empirically or by sequence analysis, using known binding sites. The regions analyzed may comprise from about 100 to about 1000 bp, e.g. from about 200, about 300, about 400, about 500 bp, up to about 1000 bp, up to about 900, up to about 800, up to about 700 bp.

[0059] Regions with EZH2 DNA-binding domains. Enhancer of Zeste 2 (EZH2) is the enzymatic subunit of Polycomb Repressive Complex 2 (PRC2), which catalyzes histone H3 lysine 27 trimethylation (H3K27me3) at target promoters for gene silencing. Regions of EZH2 DNA-binding domains may be determined empirically or by sequence analysis, using known binding sites. The regions analyzed may comprise from about 100 to about 1000 bp, e.g. from about 200, about 300, about 400, about 500 bp, up to about 1000 bp, up to about 900, up to about 800, up to about 700 bp.

[0060] Lamin-associated domains (LADs). Lamins (A/C and B) are major constituents of the nuclear lamina (NL). Structurally conserved lamina-associated domains (LADs) are formed by genomic regions that contact the NL. Lamin B1 associates with actively expressed and open

euchromatin regions, forming dynamic euchromatin lamin B1-associated domains (eLADs) of about 0.3 Mb. Chromatin regions that are in close contact with NL are called lamina-associated domains (LADs). LADs are formed by heterochromatin defined as chromatin regions with low gene frequency, transcriptionally silent, and enriched in the repressive histone marks, H3K9me2/3. Lamin associated domains may be determined empirically or by sequence analysis, using known binding sites. The regions analyzed may comprise from about 100 to about 3000 bp, e.g. from about 200, about 300, about 400, about 500 bp, up to about 3000 bp, up to about 2000, up to about 1000, up to about 750 bp.

[0061] Regions that are low methylated at birth or high methylated at birth can be defined based on a ratio of observed to expected methylation greater than or less than about 0.6. The basis may be determined with a neonatal sample or population of neonatal samples, or may be determined with an embryonic stem cell, e.g. iPSC. The regions analyzed may comprise from about 100 to about 1000 bp, e.g. from about 200, about 300, about 400, about 500 bp, up to about 1000 bp, up to about 900, up to about 800, up to about 700 bp.

[0062] Methods known in the art for determining and analyzing methylation include, inter alia, Simpson et al. *Aging Cell.* 2021 Sep;20(9):e13452; Liu Z et al. *Aging Cell.* 2020 Oct;19(10):e13229; Vijayakumar et al. *Mech Ageing Dev.* 2022 Jun;204:111676; Duan et al. *Ageing Res Rev.* 2022 Nov;81:101743; Galkin et al. *Aging Dis.* 2021 Aug 1;12(5):1252-1262; Horvath et al. *Proc Natl Acad Sci U S A.* 2022 May 24;119(21):e2120887119; Chang et al. *Front Bioinform.* 2022 Feb 10;2:815289; Di Lena et al. *Brief Bioinform.* 2022 Jul 18;23(4):bbac274. ; Ashapkin et al. *Methods Mol Biol.* 2020;2138:297-312; Amiri Roudbar et al. *G3 (Bethesda).* 2021 Jul 14;11(7):jkab112; Ryan CP *Am J Hum Biol.* 2021 May;33(3):e23488; Li et al. *PLoS Comput Biol.* 2022 Aug 19;18(8):e1009938; each herein specifically incorporated by reference.

[0063] In some embodiments, AgeIndex is used to monitor aging interventions, e.g. in the context of a clinical trial, to advise individuals, and the like. An individual may be treated in accordance with the AgeIndex analysis. For example and without limitation, interventions of interest include caloric restriction (CR) without malnutrition, e.g. caloric restriction of 10%, 15%, 20%; alone or in conjunction with diet changes. Components of a diet may include phytochemicals, essential fatty acids, etc. Specifically designed human CR experiences include CALERIE (Comprehensive Assessment of Long-term Effects of Reducing Intake of Energy) trials and the experience of members of the caloric restriction society known as Caloric Restriction with Optimal Nutrition (CRON). Genetic background, sex, percentage of CR, and the time to which CR is started have shown may be analyzed. Time-restricted feeding (TRF), intermittent fasting (IF) or fasting-mimicking diets (FMD) are among other dietetic interventions to achieve health improvement and in some cases increases in longevity.

Regularly, there is a synchronization of fasting periods with resting and repairing periods, and activity periods with energy consuming periods. This allows the coincidence of both inputs, circadian rhythm and feeding/fasting cycle, in the same direction amplifying the oscillatory patterns of expression. Interventions such as TRF or IF reinforcing the synchronization of circadian rhythms with eating behavior have been proposed as possible tools to improve health with possible influence on the life span. Fasting mimicking diets (FMD) alternate periods of several days with a diet with low calories, sugars and proteins but high in essential unsaturated fats and fiber with periods of regular diets. Thus, FMD mimics the effects of fasting, decreasing glucose and IGF-1 levels and increasing ketone bodies periodically. Changes in diet composition without significantly affecting total energy intake have also shown to present effects on health parameters and life span.

[0064] Pharmacological approaches can also be assessed. Activation of deacetylases and/or AMPK and/or inhibition of acetylases and mTORC1 pathways would lead to CR mimetic effects and several CR mimetic molecules have emerged as potential anti-aging drugs. Drugs that either stimulate processes that decay with aging (proteostasis, autophagy, mitochondrial dysfunction, etc.) or inhibit primary processes related with aging (telomere attrition, DNA instability, oxidative stress, etc.) are also potential anti-aging candidates.

[0065] The role of SIRT proteins in life extension targeted these proteins as potential sites of anti-aging interventions and triggered the search for SIRT activators, e.g. resveratrol. Other possible anti-aging molecules with CR mimetic effects are polyphenols. Among them, Curcumin has antidiabetic and cardio protective effects in rodents and caffeic acid has been shown to exert antidiabetic effects in cultured cells. Some flavonoids such as quercetin and myricetin extend life in *Caenorhabditis elegans*. Polyphenols constitute a large family of molecules acting through several mechanisms. Activation of SIRT1 and AMPK, as well as inhibition of the acetyltransferase EP300 and mTORC1, have been described to occur after polyphenol treatments. One common consequence of these signals is the activation of autophagy, which might explain health promoting effects of these molecules.

[0066] Metformin is a hypoglycemic agent and one of the most currently prescribed drugs in type 2 diabetic patients. Metformin activates AMPK as consequence of its inhibitory action on the complex I of the electron transport chain increasing the AMP/ATP ratio. Since activation of AMPK leads to the inhibition of mTORC1, metformin has emerged as a CR mimetic drug modulating two of the key effectors/mediators of CR.

[0067] Inhibitors of mTORC1 complex are also putative CR mimetic anti-aging drugs, e.g. rapamycin, which inhibits mTORC1 extending life in yeast, although rapamycin also present significant immunosuppressant effects that limit its use. This could be ameliorated using different rapamycin analogs with different pharmacokinetics and/or adjusting the dose and the schedule of its administration. A second generation of mTOR inhibitors includes new

molecules such as NVP-BEZ235, PF-04691502 and OSI-027 (TORC1/TORC2) that inhibit both mTORC1 and mTORC2, acting on mTOR kinase activity instead of targeting on FKB12. Finally, a third generation of mTOR inhibitors like RapaLink-1 (rapamycin-FRB-binding compound linked to TORKi) are bivalent molecules, binding to FRB and kinase domains.

[0068] Autophagy has been shown to improve different age-associated alterations and its benefits can be promoted by CR and a wide variety of agents such as Res, rapamycin, metformin, BRD 5611 and the polyamine spermidine.

[0069] The increase of senescent cells is one of the primary causes of the aging process. Senescent cells accumulate in the tissues of aging organisms and through its specific secretory profile, which includes the release of pro-inflammatory cytokines, proteases and several other factors, induce a low grade inflammation characteristic of age-associated chronic diseases and frailty. The combination of this low grade inflammation with aging coined the term inflammaging. A strategy to fight the aging process is the use of senolytic drugs by selectively inducing apoptosis of senescent cells. In spite of an up-regulation of pro-apoptotic pathways, senescent cells remain resistant to apoptosis due to the expression of proteins involved in different survival pathways against the pro-apoptotic profile. Targeting of these proteins has led to the discovery of the first senolytic agents, dasatinib and quercetin which interfere with ephrins receptors pathway and PI3K δ /AKT/ROS-protective pathways. New senolytic agents like Fisetin, A1331852, A1155463, Piperlongumine, Tanespimycin, Geldanmycin and Alcespimycin have come out, demonstrating their capability to reduce senescent cells with amelioration of several age-associated cardiovascular alterations and fibrosis. The combination of senolytic agents appears as a promising anti-aging intervention.

[0070] Telomerases have become another target of intervention to slow the deleterious effects of the aging process. TA-65, a telomerase activator, has been reported to increase health span in mice. In humans, a health maintenance program using the same telomerase activator have provided evidence of its beneficial effects on bone mineral density, inflammatory markers and several cardiometabolic parameters such as reduction of fasting glucose and insulin levels and lowering of circulating total and LDL cholesterol and blood pressure. AGS-499, another telomerase activator, has neuroprotective effects and delays the progression of amyotrophic lateral sclerosis in SOD1 transgenic mice.

[0071] A “dataset” is a set of numerical values resulting from evaluation of a sample (or population of samples) under a desired condition. The values of the dataset can be obtained, for example, by experimentally obtaining measures from a sample and constructing a dataset from these measurements; or alternatively, by obtaining a dataset from a service provider such as a laboratory, or from a database or a server on which the dataset has been stored. Similarly, the term “obtaining a dataset associated with a sample” encompasses obtaining a set of data

determined from at least one sample. Obtaining a dataset encompasses obtaining a sample, and processing the sample to experimentally determine the data, e.g., via measuring genomic methylation patterns. The phrase also encompasses receiving a set of data, e.g., from a third party that has processed the sample to experimentally determine the dataset.

[0072] “Measuring” or “measurement” in the context of the present teachings refers to determining the presence, absence, quantity, amount, or effective amount of methylation at specific genomic features as disclosed herein.

[0073] Classification can be made according to predictive modeling methods that set a threshold for determining the probability that a sample belongs to a given class. The probability preferably is at least 50%, or at least 60% or at least 70% or at least 80% or higher. Classifications also can be made by determining whether a comparison between an obtained dataset and a reference dataset yields a statistically significant difference. If so, then the sample from which the dataset was obtained is classified as not belonging to the reference dataset class. Conversely, if such a comparison is not statistically significantly different from the reference dataset, then the sample from which the dataset was obtained is classified as belonging to the reference dataset class.

[0074] Classification is the process of recognizing, understanding, and grouping ideas and objects into preset categories or “sub-populations.” Using pre-categorized training datasets, machine learning programs use a variety of algorithms to classify future datasets into categories. Classification algorithms in machine learning use input training data to predict the likelihood that subsequent data will fall into one of the predetermined categories. An analytic classification process may use any one of a variety of statistical analytic methods to manipulate the quantitative data and provide for classification of the sample. Examples of useful methods include linear discriminant analysis, recursive feature elimination, a prediction analysis of microarray, a logistic regression, a CART algorithm, a FlexTree algorithm, a LART algorithm, a random forest algorithm, a MART algorithm, machine learning algorithms; etc. Using any one of these methods, a protein distribution pattern may be used to generate a predictive model. In the generation of such a model, a dataset comprising differently aged cells are used as a training set. A training set will contain data for one or more different distributions of interest. In some embodiments a decision tree is used to order classes on a precise level, for example with a random forest algorithm.

[0075] The predictive ability of a model can be evaluated according to its ability to provide a quality metric, e.g. AUC or accuracy, of a particular value, or range of values. In some embodiments, a desired quality threshold is a predictive model that will classify a sample with an accuracy of at least about 0.7, at least about 0.75, at least about 0.8, at least about 0.85, at least about 0.9, at least about 0.95, or higher. As an alternative measure, a desired quality threshold can refer to a predictive model that will classify a sample with an AUC (area under

the curve) of at least about 0.7, at least about 0.75, at least about 0.8, at least about 0.85, at least about 0.9, or higher.

[0076] As is known in the art, the relative sensitivity and specificity of a predictive model can be “tuned” to favor either the selectivity metric or the sensitivity metric, where the two metrics have an inverse relationship. The limits in a model as described above can be adjusted to provide a selected sensitivity or specificity level, depending on the particular requirements of the test being performed. One or both of sensitivity and specificity can be at least about at least about 0.7, at least about 0.75, at least about 0.8, at least about 0.85, at least about 0.9, or higher.

[0077] The raw data may be initially analyzed by measuring the values for each marker, usually in triplicate or in multiple triplicates; and the cells may be clustered into populations, e.g. with flowSOM. The data may be manipulated, for example, raw data may be transformed using standard curves, and the average of triplicate measurements used to calculate the average and standard deviation for each patient. These values may be transformed before being used in the models, e.g. log-transformed, Box-Cox transformed (see Box and Cox (1964) J. Royal Stat. Soc., Series B, 26:211–246), etc. The data are then input into a predictive model, which will classify the sample according to the state. The resulting information may be transmitted to a patient or health professional.

[0078] Analysis of biological samples, e.g. cell samples, circulating DNA samples, etc., obtained from an individual is used to obtain a determination of changes in methylation status, which are shown herein to be predictive of molecular aging. The sample can be any suitable type that allows for such analysis. Samples can be obtained once or multiple times from an individual. The cells can be separated from body samples by red cell lysis, centrifugation, elutriation, density gradient separation, apheresis, affinity selection, panning, FACS, centrifugation with Hypaque, solid supports (magnetic beads, beads in columns, or other surfaces) with attached antibodies, etc.

[0079] A phenotypic profile of a DNA population factors in the presence of methylation markers at defined genomic features. It is understood that marker levels can exist as a distribution and that a marker used to classify a cell can be a particular point on the distribution but more typically can be a portion of the distribution.

[0080] Samples may be obtained at one or more time points. Where a sample at a single time point is used, comparison is made to a reference “base line” level for the feature, which may be obtained from a training set data as disclosed herein. Samples can also be obtained over time for a subject, e.g. before and after cancer therapy, aging intervention, and the like; or over chronologic time. For example, samples can be obtained at one, two, three, four or more

time points. The time points may be before and after a therapy of interest, e.g. a cancer treatment; aging intervention therapy, and the like.

[0081] In some embodiment, the methods of the invention include the use of liquid handling components. The liquid handling systems can include robotic systems comprising any number of components. In addition, any or all of the steps outlined herein can be automated; thus, for example, the systems can be completely or partially automated. As will be appreciated by those in the art, there are a wide variety of components which can be used, including, but not limited to, one or more robotic arms; plate handlers for the positioning of microplates; automated lid or cap handlers to remove and replace lids for wells on non-cross contamination plates; tip assemblies for sample distribution with disposable tips; washable tip assemblies for sample distribution; 96 well loading blocks; cooled reagent racks; microtiter plate pipette positions (optionally cooled); stacking towers for plates and tips; and computer systems.

[0082] Fully robotic or microfluidic systems include automated liquid-, particle-, cell- and organism-handling including high throughput pipetting to perform all steps of screening applications. This includes liquid, particle, cell, and organism manipulations such as aspiration, dispensing, mixing, diluting, washing, accurate volumetric transfers; retrieving, and discarding of pipet tips; and repetitive pipetting of identical volumes for multiple deliveries from a single sample aspiration. These manipulations are cross-contamination- free liquid, particle, cell, and organism transfers. This instrument performs automated replication of microplate samples to filters, membranes, and/or daughter plates, high-density transfers, full-plate serial dilutions, and high capacity operation.

[0083] In some embodiments, platforms for multi-well plates, multi-tubes, holders, cartridges, minitubes, deep-well plates, microfuge tubes, cryovials, square well plates, filters, chips, optic fibers, beads, and other solid-phase matrices or platform with various volumes are accommodated on an upgradable modular platform for additional capacity. This modular platform includes a variable speed orbital shaker, and multi-position work decks for source samples, sample and reagent dilution, assay plates, sample and reagent reservoirs, pipette tips, and an active wash station. In some embodiments, the methods of the invention include the use of a plate reader.

[0084] In some embodiments, interchangeable pipet heads (single or multi-channel) with single or multiple magnetic probes, affinity probes, or pipettors robotically manipulate the liquid, particles, cells, and organisms. Multi-well or multi-tube magnetic separators or platforms manipulate liquid, particles, cells, and organisms in single or multiple sample formats.

[0085] In some embodiments, the instrumentation will include a detector, which can be a wide variety of different detectors, depending on the labels and assay. In some embodiments, useful detectors include a mass cyometer; and a computer workstation.

[0086] In some embodiments, the robotic apparatus includes a central processing unit which communicates with a memory and a set of input/output devices (e.g., keyboard, mouse, monitor, printer, etc.) through a bus. Again, as outlined below, this can be in addition to or in place of the CPU for the multiplexing devices of the invention. The general interaction between a central processing unit, a memory, input/output devices, and a bus is known in the art. Thus, a variety of different procedures, depending on the experiments to be run, are stored in the CPU memory.

Data Analysis

[0087] An AgeIndex can be generated from a biological sample using any convenient protocol, for example as described herein. The readout can be a mean, average, median or the variance or other statistically or mathematically-derived value associated with the measurement. The marker readout information can be further refined by direct comparison with the corresponding reference or control pattern. A population distribution pattern can be evaluated on a number of points: to determine if there is a statistically significant change at any point in the data matrix relative to a reference value; whether the change is an increase or decrease in the population frequency; and the like. The absolute values obtained for each marker under identical conditions will display a variability that is inherent in live biological systems and also reflects the variability inherent between individuals.

[0088] Following obtainment of the AgeIndex from the sample being assayed, the AgeIndex can be compared with a reference or base line profile to make a prognosis or analysis regarding the phenotype of the patient from which the sample was obtained/derived.

[0089] In certain embodiments, the obtained AgeIndex is compared to a single reference/control profile to obtain information regarding the aging phenotype of the patient being assayed. In yet other embodiments, the obtained AgeIndex is compared to two or more different reference/control profiles to obtain more in depth information regarding the phenotype of the patient. For example, the obtained AgeIndex can be compared to a young and aged reference to obtain confirmed information regarding whether the patient has the phenotype of interest.

[0090] The data can be subjected to non-supervised hierarchical clustering to reveal relationships among profiles. For example, hierarchical clustering can be performed, where the Pearson correlation is employed as the clustering metric. One approach is to consider a patient disease dataset as a "learning sample" in a problem of "supervised learning". CART is a standard in applications to medicine (Singer (1999) Recursive Partitioning in the Health Sciences, Springer), which can be modified by transforming any qualitative features to quantitative features; sorting them by attained significance levels, evaluated by sample reuse methods for Hotelling's T^2 statistic; and suitable application of the lasso method. Problems in

prediction are turned into problems in regression without losing sight of prediction, indeed by making suitable use of the Gini criterion for classification in evaluating the quality of regressions.

[0091] Other methods of analysis that can be used include logistic regression. One method of logic regression Ruczinski (2003) *Journal of Computational and Graphical Statistics* 12:475-512. Logic regression resembles CART in that its classifier can be displayed as a binary tree. It is different in that each node has Boolean statements about features that are more general than the simple “and” statements produced by CART.

[0092] Another approach is that of nearest shrunken centroids (Tibshirani (2002) *PNAS* 99:6567-72). The technology is k-means-like, but has the advantage that by shrinking cluster centers, one automatically selects features (as in the lasso) so as to focus attention on small numbers of those that are informative. The approach is available as Prediction Analysis of Microarrays (PAM) software, a software “plug-in” for Microsoft Excel, and is widely used. Two further sets of algorithms are random forests (Breiman (2001) *Machine Learning* 45:5-32 and MART (Hastie (2001) *The Elements of Statistical Learning*, Springer). These two methods are already “committee methods.” Thus, they involve predictors that “vote” on outcome. Several of these methods are based on the “R” software, developed at Stanford University, which provides a statistical framework that is continuously being improved and updated in an ongoing basis.

[0093] Other statistical analysis approaches including principle components analysis, recursive partitioning, predictive algorithms, Bayesian networks, random forest, and neural networks.

[0094] The analysis and database storage can be implemented in hardware or software, or a combination of both. In one embodiment of the invention, a machine-readable storage medium is provided, the medium comprising a data storage material encoded with machine readable data which, when using a machine programmed with instructions for using said data, is capable of displaying a any of the datasets and data comparisons of this invention. Such data can be used for a variety of purposes, such as patient monitoring, initial diagnosis, clinical trial analysis, and the like. Preferably, the invention is implemented in computer programs executing on programmable computers, comprising a processor, a data storage system (including volatile and non-volatile memory and/or storage elements), at least one input device, and at least one output device. Program code is applied to input data to perform the functions described above and generate output information. The output information is applied to one or more output devices, in known fashion. The computer can be, for example, a personal computer, microcomputer, or workstation of conventional design.

[0095] Each program is preferably implemented in a high level procedural or object oriented programming language to communicate with a computer system. However, the programs can

be implemented in assembly or machine language, if desired. In any case, the language can be a compiled or interpreted language. Each such computer program is preferably stored on a storage media or device readable by a general or special purpose programmable computer, for configuring and operating the computer when the storage media or device is read by the computer to perform the procedures described herein. The system can also be considered to be implemented as a computer-readable storage medium, configured with a computer program, where the storage medium so configured causes a computer to operate in a specific and predefined manner to perform the functions described herein.

[0096] A variety of structural formats for the input and output means can be used to input and output the information in the computer-based systems of the present invention. One format for an output means test datasets possessing varying degrees of similarity to a trusted profile. Such presentation provides a skilled artisan with a ranking of similarities and identifies the degree of similarity contained in the test pattern.

[0097] The signature patterns and databases thereof can be provided in a variety of media to facilitate their use. "Media" refers to a manufacture that contains the signature pattern information of the present invention. The databases of the present invention can be recorded on computer readable media, *e.g.* any medium that can be read and accessed directly by a computer. Such media include, but are not limited to: magnetic storage media, such as floppy discs, hard disc storage medium, and magnetic tape; optical storage media such as CD-ROM; electrical storage media such as RAM and ROM; and hybrids of these categories such as magnetic/optical storage media. One of skill in the art can readily appreciate how any of the presently known computer readable mediums can be used to create a manufacture comprising a recording of the present database information. "Recorded" refers to a process for storing information on computer readable medium, using any such methods as known in the art. Any convenient data storage structure can be chosen, based on the means used to access the stored information. A variety of data processor programs and formats can be used for storage, *e.g.* word processing text file, database format, *etc.*

Kits

[0098] In some embodiments, the invention provides kits for the classification, diagnosis, prognosis, theragnosis, and/or prediction of molecular aging in a subject. The kit may further comprise a software package for data analysis of the cellular state and its physiological status, which may include reference profiles for comparison with the test profile and comparisons to other analyses as referred to above. The kit may also include instructions for use for any of the above applications.

[0099] Such kits may also include information, such as scientific literature references, package insert materials, clinical trial results, and/or summaries of these and the like, which

indicate or establish the activities and/or advantages of the composition, and/or which describe dosing, administration, side effects, drug interactions, or other information useful to the health care provider. Such information may be based on the results of various studies, for example, studies using experimental animals involving in vivo models and studies based on human clinical trials. Kits described herein can be provided, marketed and/or promoted to health providers, including physicians, nurses, pharmacists, formulary officials, and the like. Kits may also, in some embodiments, be marketed directly to the consumer.

Reports

[00100] In some embodiments, providing an evaluation of a subject for a classification, diagnosis, prognosis, theranosis, and/or prediction of biological aging includes generating a written report that includes the artisan's assessment of the subject's state of health, including, for example, a "diagnosis assessment", of the subject's prognosis, i.e. a "prognosis assessment", and/or of possible treatment regimens, i.e. a "treatment assessment". Thus, a subject method may further include a step of generating or outputting a report providing the results of an assessment, which report can be provided in the form of an electronic medium (e.g., an electronic display on a computer monitor), or in the form of a tangible medium (e.g., a report printed on paper or other tangible medium).

[00101] A "report," as described herein, is an electronic or tangible document which includes report elements that provide information of interest relating to a diagnosis assessment, a prognosis assessment, and/or a treatment assessment and its results. A subject report can be completely or partially electronically generated. A subject report includes at least a diagnosis assessment, and/or a suggested course of treatment to be followed. A subject report can further include one or more of: 1) information regarding the testing facility; 2) service provider information; 3) subject data; 4) sample data; 5) an assessment report, which can include various information including: a) test data, where test data can include an analysis of cellular signaling responses to activation, b) reference values employed, if any.

[00102] The report may include information about the testing facility, which information is relevant to the hospital, clinic, or laboratory in which sample gathering and/or data generation was conducted. This information can include one or more details relating to, for example, the name and location of the testing facility, the identity of the lab technician who conducted the assay and/or who entered the input data, the date and time the assay was conducted and/or analyzed, the location where the sample and/or result data is stored, the lot number of the reagents (e.g., kit, etc.) used in the assay, and the like. Report fields with this information can generally be populated using information provided by the user.

[00103] The report may include information about the service provider, which may be located outside the healthcare facility at which the user is located, or within the healthcare facility.

Examples of such information can include the name and location of the service provider, the name of the reviewer, and where necessary or desired the name of the individual who conducted sample gathering and/or data generation. Report fields with this information can generally be populated using data entered by the user, which can be selected from among pre-scripted selections (e.g., using a drop-down menu). Other service provider information in the report can include contact information for technical information about the result and/or about the interpretive report.

[00104] The report may include a subject data section, including subject medical history as well as administrative subject data (that is, data that are not essential to the diagnosis, prognosis, or treatment assessment) such as information to identify the subject (e.g., name, subject date of birth (DOB), gender, mailing and/or residence address, medical record number (MRN), room and/or bed number in a healthcare facility), insurance information, and the like), the name of the subject's physician or other health professional who ordered the susceptibility prediction and, if different from the ordering physician, the name of a staff physician who is responsible for the subject's care (e.g., primary care physician).

[00105] The report may include a sample data section, which may provide information about the biological sample analyzed, such as the source of biological sample obtained from the subject (e.g. blood, type of tissue, etc.), how the sample was handled (e.g. storage temperature, preparatory protocols) and the date and time collected. Report fields with this information can generally be populated using data entered by the user, some of which may be provided as pre-scripted selections (e.g., using a drop-down menu).

[00106] The report may include an assessment report section, which may include information generated after processing of the data as described herein. The interpretive report can include a prognosis of the likelihood that the patient will develop preeclampsia. The interpretive report can include, for example, results of the analysis, methods used to calculate the analysis, and interpretation, i.e. prognosis. The assessment portion of the report can optionally also include a Recommendation(s).

[00107] It will also be readily appreciated that the reports can include additional elements or modified elements. For example, where electronic, the report can contain hyperlinks which point to internal or external databases which provide more detailed information about selected elements of the report. For example, the patient data element of the report can include a hyperlink to an electronic patient record, or a site for accessing such a patient record, which patient record is maintained in a confidential database. This latter embodiment may be of interest in an in-hospital system or in-clinic setting. When in electronic format, the report is recorded on a suitable physical medium, such as a computer readable medium, e.g., in a computer memory, zip drive, CD, DVD, etc.

[00108] It will be readily appreciated that the report can include all or some of the elements above, with the proviso that the report generally includes at least the elements sufficient to provide the analysis requested by the user (e.g., a diagnosis, a prognosis, or a prediction of responsiveness to a therapy).

EXPERIMENTAL

[00109] The following examples are put forth so as to provide those of ordinary skill in the art with a complete disclosure and description of how to make and use the present invention, and are not intended to limit the scope of what the inventors regard as their invention nor are they intended to represent that the experiments below are all or the only experiments performed. Efforts have been made to ensure accuracy with respect to numbers used (e.g. amounts, temperature, etc.) but some experimental errors and deviations should be accounted for. Unless indicated otherwise, parts are parts by weight, molecular weight is weight average molecular weight, temperature is in degrees Centigrade, and pressure is at or near atmospheric.

EXAMPLE 1

AgeIndex, a whole-genome epigenetic aging and rejuvenation index

[00110] DNA methylation (DNAm) is one of the epigenetic marks most strongly associated with aging. While there is increasing evidence that this association is driven mainly by the age-dependent global loss of epigenetic information, existing approaches have focused exclusively on selecting CpGs that are most correlated with age, without regard to underlying genomic features. We have identified herein genomic features associated with age-dependent modulation of DNA methylation, and developed a novel measure of epigenetic information loss, AgeIndex, which captures bidirectional transitions in genome-wide methylation states. These transitions reflect methylation gain in unmethylated, CpG-dense regions enriched for EZH2 or EHMT2 binding, and methylation loss in methylated, CpG-sparse regions associated with lamin. In contrast to methylation clocks and epigenome-wide association studies (EWAS), AgeIndex is assay-agnostic and can be trained with small numbers of samples. We also demonstrate it is more suitable for studying aging interventions.

[00111] *Genomic features associated with DNAm at birth.* To identify genomic and epigenomic features associated with DNA methylation levels at birth, we analyzed a large DNA methylation dataset from 438 neonatal whole blood samples (GSE103657), assayed using Illumina Infinium 450K BeadChip (Methyl450). To understand the impact of DNA sequence context on CpG methylation, we computed densities for all 16 possible dinucleotides within a 1kb window flanking each CpG. In agreement with previous studies, we observed that DNA methylation

was most negatively correlated with CpG density. We also made the novel observation that DNA methylation was most positively correlated with TpG and CpA density (Fig 1a). The theory of “hypodeamination” proposes that because methylated CpGs are more prone to spontaneous deamination into cytosine than unmethylated CpGs, unmethylated CpG-dense regions can sustain higher CpG content throughout evolution, making them less prone to mutation. Hypodeamination elegantly explains the negative correlation between CpG density and methylation, and may also explain the positive correlation of methylation with TpG and CpA density. This theory motivated us to design a new measure of “CpG-protection,” defined as CpG density divided by TpG+CpA density, that better reflects this evolutionary force. Comparing regions annotated as CpG-protected with regions annotated as CpG islands identified a significantly stronger association with hypomethylation for CpG-protected regions (Fig 1b).

[00112] To further understand the effects of genomic context on methylation, we also compared DNA methylation at the DNA-binding regions for all transcription factors (TFs) from ENCODE3 datasets with chromatin immunoprecipitation sequencing (ChIP-Seq) assayed in at least three cell types (121 TFs in total). Regions with DNA-binding domains for EZH2 methyltransferase had the highest DNA methylation levels among all TFs when controlling for CpG density. Regions with DNA-binding for EHMT2 methyltransferase also displayed a high level of DNA methylation (Fig 1c).

[00113] For genomic features with causal effects on modulating DNA methylation, time might be a moderator. To examine if any of the identified features were associated with changes in DNA methylation in a time-dependent manner, we compared DNA methylation levels between neonates and adults. We analyzed a large methylation dataset from 799 adult (age range) whole blood samples (GSE152027), assayed using the Methyl450 platform, and processed using the same pipeline as the neonatal dataset. As expected, we observed a global loss of DNA methylation at lamin-associated CpG-sparse regions ([Zhou et al., 2018](#)). We also observed a global gain of DNA methylation in CpG-rich regions, as defined by our CpG-protection measure (Fig 1d). Similar effects were observed for EZH2/EHMT2 DNA-binding domains. To quantify the marginal contribution of each genomic feature to age-dependent changes in methylation, we performed a regression analysis, with DNA methylation levels in adults as the response and all the genomic features as predictors. Controlling for initial DNA methylation state at birth, CpG protection was identified as the feature with the highest positive marginal effect (p-value), followed by EZH2 and EHMT2 bindings (p-values). Lamin was associated with a significant negative marginal effect on age-dependent DNA methylation (Fig 1e).

[00114] *Genomic features associated with DNAm with aging.* The observed differential effect of genomic features on DNA methylation between neonates and adults could be limited to a

specific age range (e.g. childhood) or holds across the lifespan. To examine the effect of each feature on age-dependent changes in DNA methylation throughout the human lifespan, we analyzed Methyl450 DNA methylation data from 4,518 blood samples of different ages (13 to 100 years old) from four large DNA methylation datasets (GSE152027, GSE40279, GSE42861, and GSE55763). This data confirmed that lamin-associated CpG-poor regions, identified in [Zhou et al. 2018](#), significantly lose DNA methylation with age (p-values, Fig 2a). We also observed that highly methylated regions at birth as well as regions with high density of TpG and CpA dinucleotides lose methylation across lifespan (p-values). Conversely, DNA methylation levels of lowly methylated regions at birth as well CpG-rich regions and DNA-binding regions for EZH2 and EHMT2 all increase with age, across all datasets (p-values). These age-dependent changes for each genomic feature are summarized in Fig 2.

[00115] Examining the 7 genomic features we identified, we observed that all features associated with an age-dependent decrease in DNA methylation corresponded to highly methylated regions at birth, while all features with increasing DNA methylation corresponded to lowly methylated regions. Therefore, the data suggest an age-dependent transition of the epigenome from initial low or high states at birth toward a partially methylated state later in life. This observation provides empirical evidence for “loss of epigenetic information” with age.

[00116] *AgeIndex, a feature-based biomarker of aging.* The features we identified as associated with the age-dependent changes in the epigenome are combined to capture the epigenetic state of cells in a more holistic approach compared to epigenetic clocks or EWAS studies. Leveraging on this evidence, we integrated all this information in an algorithm that allowed us to develop AgeIndex, a machine-learning based approach, able to measure the epigenetic loss of information by capturing genome-wide gain or loss of DNA methylation occurring during age. We first divided genomic features into two groups based on their association with the gain or loss of DNA methylation. We then aggregate DNA methylation levels across regions associated with genomic features in each group, resulting in AgeIndex_gain and AgeIndex_loss indices. The large numbers of CpG sites in each group makes AgeIndex an assay-agnostic method, and allows us to aggregate changes in methylation across different types of genomic regions. For example, we can perform a metagene analysis to study the loss and gain of epigenetic information around the transcription start site (TSS) of genes, or even identify changes at the level of individual chromosomes. We applied AgeIndex to multiple large blood methylation datasets assayed using Methyl450, Infinium MethylationEPIC (Epic), Methyl-CpG-Binding Domain Sequencing (MBD-Seq, [McClay 2014](#)), and WGBS (mainly from the BLUEPRINT epigenome project, [Martens 2013](#)). Across all of these assays, we observed a very consistent pattern reflecting global loss of epigenetic information (Fig 3a), which appears to be shared across chromosomes (Fig 3b).

These trends were also consistent across multiple cell types, indicating that this information loss is not cell type-specific (Fig 3c).

Materials and Methods

- [00117] *WG Methyl-Seq – Library Preparation and Illumina Sequencing.* Library preparations, sequencing reactions, and data analysis were conducted at Azenta Life Sciences (South Plainfield, NJ, USA). DNA samples were quantified using Qubit 2.0 Fluorometer (Life Technologies, Carlsbad, CA, USA).
- [00118] *Library preparation.* 100 ng gDNA was combined with 0.001 ng CpG methylated pUC19 and 0.02 ng unmethylated lambda control DNA and sheared using Covaris LE220 Focused-ultrasonicator for an average 350 bp size. The sheared material was transferred to a PCR strip tube to begin library construction. NEBNext DNA Ultra II Reagents (NEB, Ipswich, MA) were used according to the manufacturer's instructions for end repair, A-tailing and adaptor ligation of EM-seq adaptor. The ligated samples were cleaned up according to the manufacturer's instructions with NEBNext Sample Purification Beads.
- [00119] Ligated DNA was oxidized by TET2 enzymatic reaction, which was initiated by adding Fe (II) solution and then incubated for 1 h at 37 °C. Following this, Stop Reagent was added and continue incubated for 30 minutes at 37 °C.
- [00120] Oxidized DNA was cleaned up according to the manufacturer's instructions with NEBNext Sample Purification Bead, denatured by Formamide, and deaminated with APOBEC enzymatic reaction, which was incubated at 37 °C for 3 hours.
- [00121] Deaminated DNA was cleaned up according to the manufacturer's instructions with NEBNext Sample Purification Bead, then PCR amplified with NEBNext Q5U Master Mix and EM-Seq Index Primers.
- [00122] Amplified Library was cleaned up according to the manufacturer's instructions with NEBNext Sample Purification Bead. The final library was assessed with Qubit 4.0 Fluorometer and Agilent TapeStation, and final quantified by qPCR.
- [00123] *Illumina sequencing.* The sequencing libraries were multiplexed and sequenced on the Illumina HiSeq instrument (4000 or equivalent) according to manufacturer's instructions. The samples were sequenced using a 2x150 Paired End (PE) configuration.
- [00124] *Bioinformatics Analysis.* Sequence data was trimmed to remove possible adapter sequences and nucleotides with poor quality before analysis using Illumina HiSeq analysis software v2.1 (HAS2.1). Trimmed sequence reads were aligned to the reference genomes for homo sapiens (NCBI GRCh38) with the Bismark v.0.18.1. Followed by alignment, Duplicated reads were identified and marked with Bismark deduplicate tool. These reads were removed from downstream analysis. Bismark Methylation Extractor was used to call methylation status of individual bases in the deduplicated aligned reads.

References

- [00125] Shchukina, I., Bagaitkar, J., Shpynov, O., Loginicheva, E., Porter, S., Mogilenko, D. A., ... & Artyomov, M. N. (2020). Epigenetic aging of classical monocytes from healthy individuals. *BioRxiv*.
- [00126] Vandiver, A. R., Irizarry, R. A., Hansen, K. D., Garza, L. A., Runarsson, A., Li, X., ... & Feinberg, A. P. (2015). Age and sun exposure-related widespread genomic blocks of hypomethylation in nonmalignant skin. *Genome biology*, 16(1), 1-15.
- [00127] Vandiver, Amy R., et al. "Age and sun exposure-related widespread genomic blocks of hypomethylation in nonmalignant skin." *Genome biology* 16.1 (2015): 1-15. Age and sun exposure-related widespread genomic blocks of hypomethylation in nonmalignant skin - Genome Biologyn loss in late-replicating domains is linked to mitotic cell division.
- [00128] Jenkinson, G., Pujadas, E., Goutsias, J., & Feinberg, A. P. (2017). Potential energy landscapes identify the information-theoretic nature of the epigenome. *Nature genetics*, 49(5), 719-729.
- [00129] Zhou, et al. (2018). DNA methylation loss in late-replicating domains is linked to mitotic cell division. *Nature genetics*, 50(4), 591-602.
- [00130] Kerepesi, C., Zhang, B., Lee, S. G., Trapp, A., & Gladyshev, V. N. (2021). Epigenetic clocks reveal a rejuvenation event during embryogenesis followed by aging. *bioRxiv*.
- [00131] Ahadi, Sara, et al. "Personal aging markers and ageotypes revealed by deep longitudinal profiling." *Nature medicine* 26.1 (2020): 83-90. Personal aging markers and ageotypes revealed by deep longitudinal profiling | Nature Medicine
- [00132] Cohen, Netta Mendelson, Ephraim Kenigsberg, and Amos Tanay. "Primate CpG islands are maintained by heterogeneous evolutionary regimes involving minimal selection." *Cell* 145.5 (2011): 773-786.
- [00133] McClay, Joseph L., et al. "A methylome-wide study of aging using massively parallel sequencing of the methyl-CpG-enriched genomic fraction from blood in over 700 subjects." *Human molecular genetics* 23.5 (2014): 1175-1185.
- [00134] Martens, Joost HA, and Hendrik G. Stunnenberg. "BLUEPRINT: mapping human blood cell epigenomes." *Haematologica* 98.10 (2013): 1487.
- [00135] The preceding merely illustrates the principles of the invention. It will be appreciated that those skilled in the art will be able to devise various arrangements which, although not explicitly described or shown herein, embody the principles of the invention and are included within its spirit and scope. Furthermore, all examples and conditional language recited herein are principally intended to aid the reader in understanding the principles of the invention and the concepts contributed by the inventors to furthering the art, and are to be construed as

being without limitation to such specifically recited examples and conditions. Moreover, all statements herein reciting principles, aspects, and embodiments of the invention as well as specific examples thereof, are intended to encompass both structural and functional equivalents thereof. Additionally, it is intended that such equivalents include both currently known equivalents and equivalents developed in the future, i.e., any elements developed that perform the same function, regardless of structure. The scope of the present invention, therefore, is not intended to be limited to the exemplary embodiments shown and described herein. Rather, the scope and spirit of the present invention is embodied by the appended claims.

THAT WHICH IS CLAIMED IS:

1. A method for determining a whole genome epigenetic aging index for an individual, the method comprising:

determining the methylation status of features of a genome or fraction thereof in a DNA sample;

grouping genomic feature data by association with the gain or loss of DNA methylation, where age-associated loss of methylation is found with (a) high methylated regions at birth; and (b) lamin-associated regions; and age-associated gain of methylation is observed with (a) CpG-protected regions; (b) EZH2 DNA-binding domains; (c) EHMT2 DNA-binding domains; and (d) low-methylated regions at birth;

aggregating DNA methylation levels across the features in the two groups to provide AgeIndex_gain and AgeIndex_loss indices;

integrating the AgeIndex_gain and AgeIndex_loss indices in a machine learning algorithm that compiles the epigenetic loss of information in an AgeIndex.

2. The method of claim 1, wherein the methylation status is determined for a whole genome.

3. The method of claim 1, wherein the methylation status is determined for an individual chromosome.

4. The method of any of claims 1-3, wherein the DNA sample is obtained from a cellular sample of the individual.

5. The method of any of claims 1-3, wherein the DNA sample is a circulating DNA sample from the individual.

6. The method of any of claims 1-5, wherein the methylation status is determined by bisulfite conversion followed by sequencing.

7. The method of any of claims 1-6, wherein the AgeIndex performs a metagene analysis of loss and gain of epigenetic information around the transcription start site (TSS) of genes.

8. The method of any of claims 1-6, wherein the AgeIndex performs a metagene analysis of loss and gain of epigenetic information in specific tissues.

9. The method of any of claims 1-6, wherein the AgeIndex performs a metagene analysis of loss and gain of epigenetic information in cancer cells.

10. The method of any of claims 1-6, wherein multiple AgeIndex analyses are performed on an individual over time.

11. The method of claim 10, wherein the individual is assessed before and after a therapy of interest.

12. The method of claim 11, wherein the therapy of interest is a cancer treatment.

13. The method of claim 11, wherein the therapy of interest is aging intervention therapy.

14. The method of any of claims 1-13, wherein the method is executed through the use of a computer based software program wherein the methylation features are input and the software program outputs a score indicative of a particular classification as defined by the user.

15. The method of claim 14, wherein the software programs employs machine learning to uncover relationships between input metrics in their relation to target outputs through training algorithms.

16. The method of any of claims 1-15, further comprising providing a report to the individual with an AgeIndex assessment.

Fig. 1: Genomic features associate with DNA methylation at birth

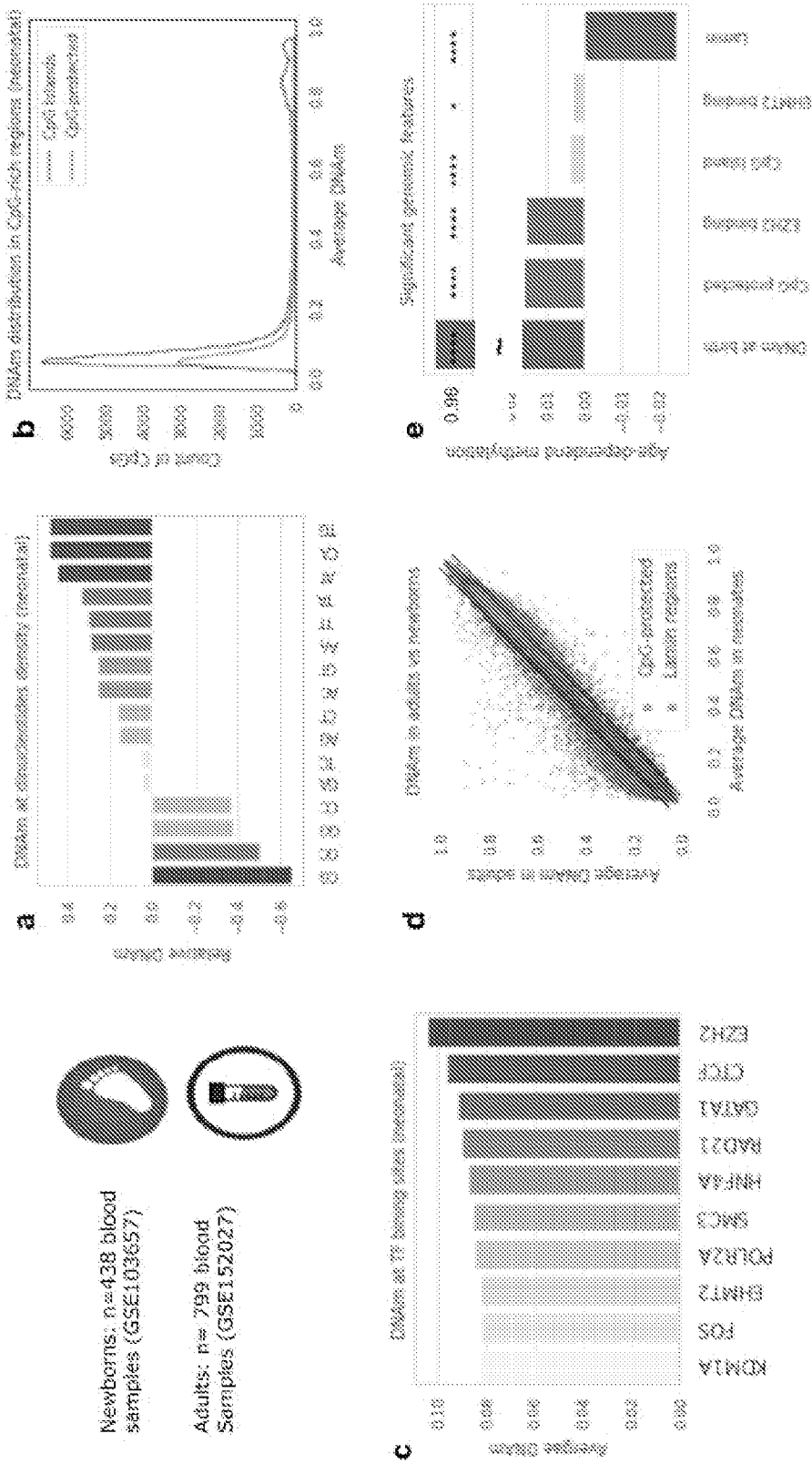


Fig. 2: Genomic features associates with age-dependent DNA methylation

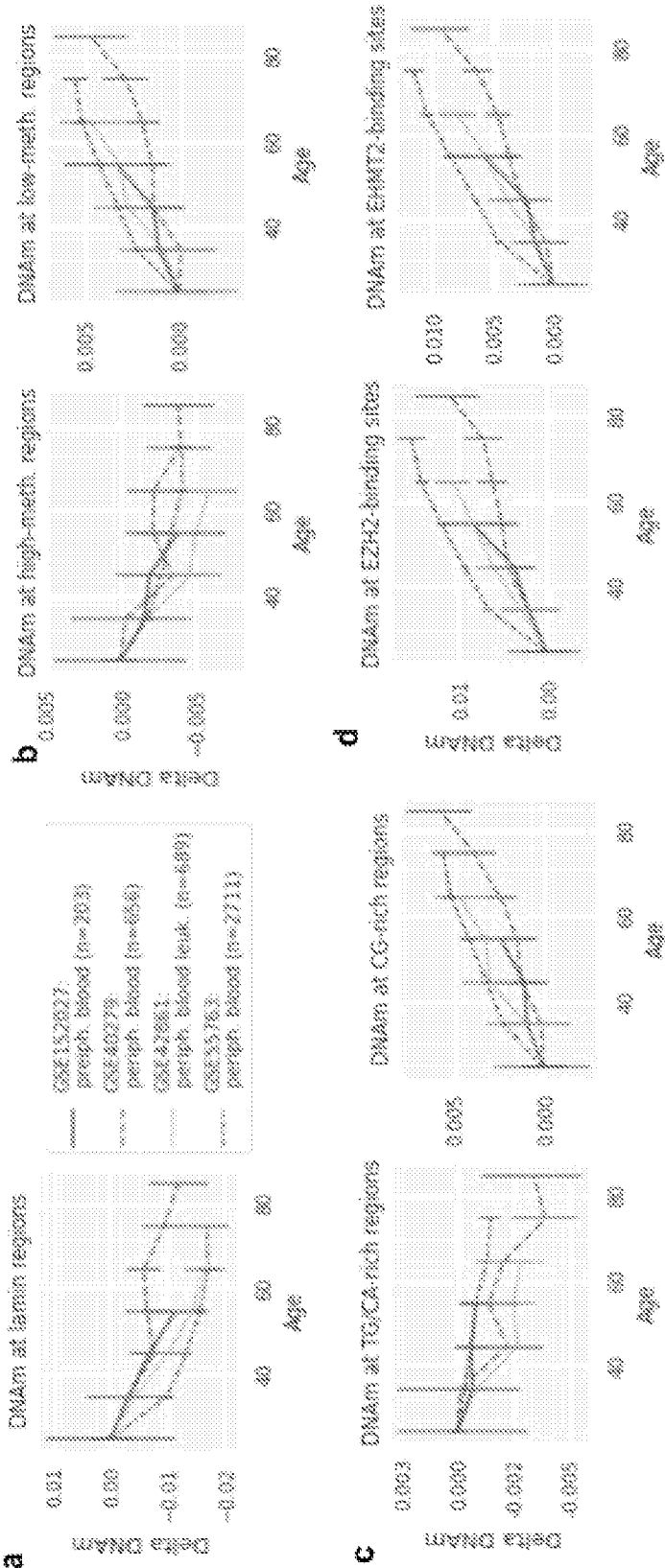
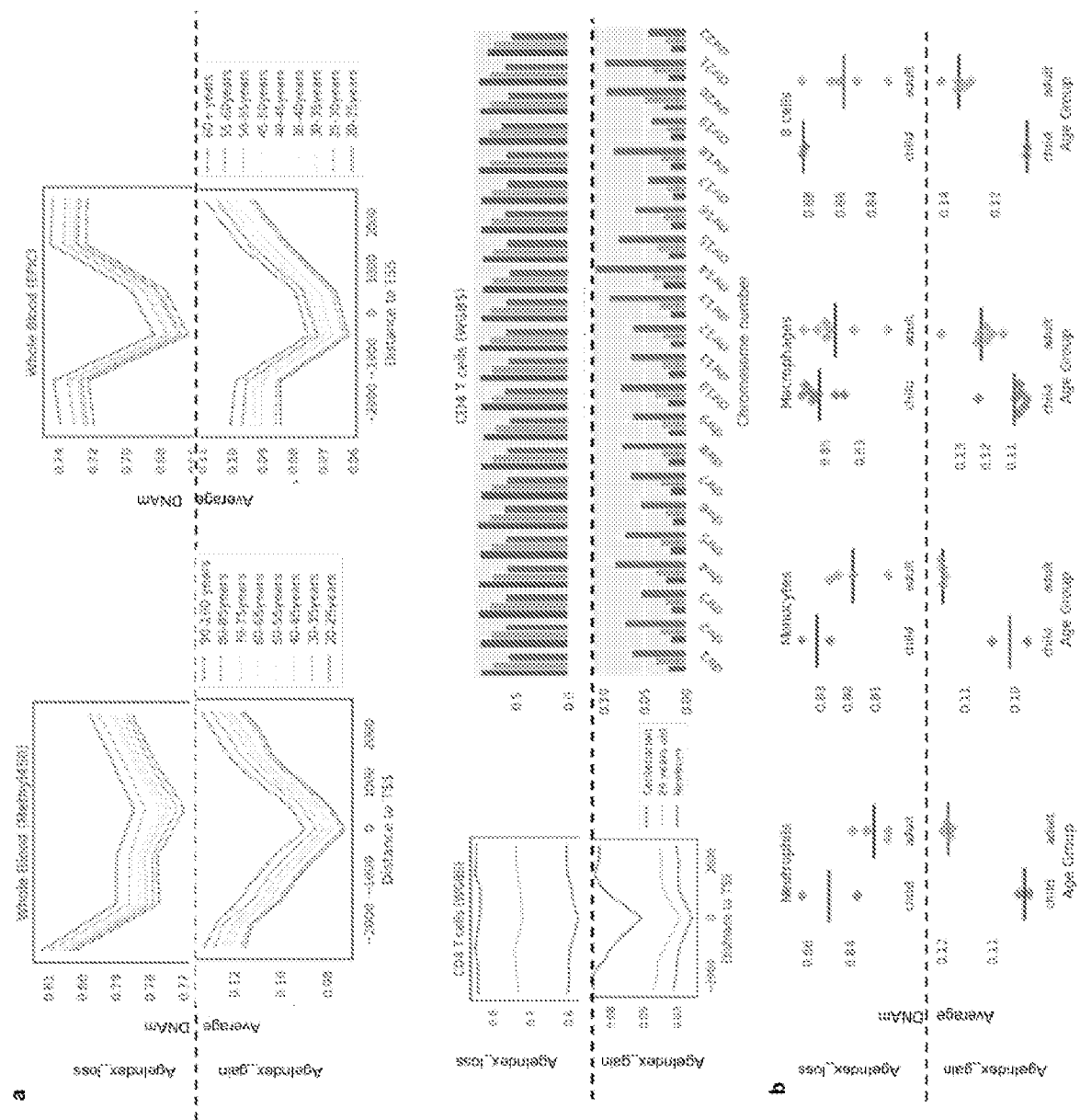


Fig. 3: Ageindex differentiates sorted and whole blood from different age groups, assayed by different methods



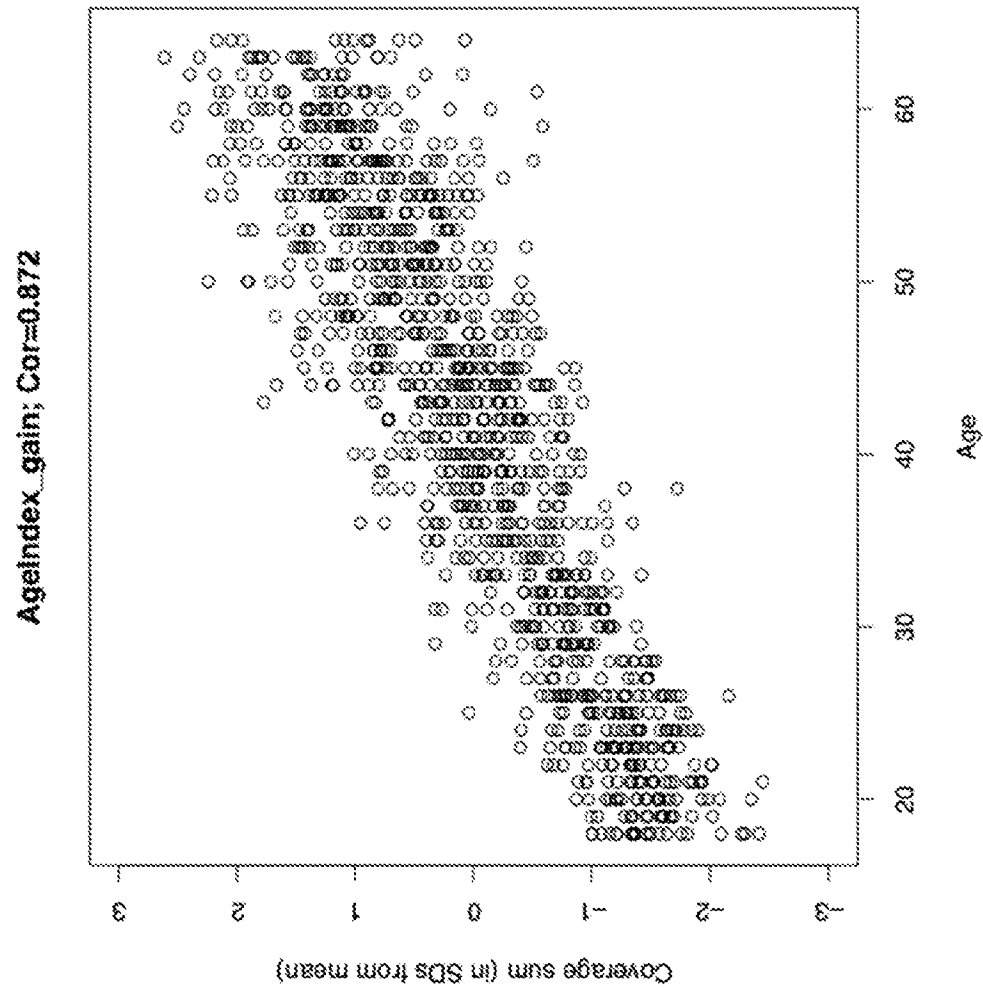


FIG. 4

FIG. 5
AgeIndex associates with aging in different tissues

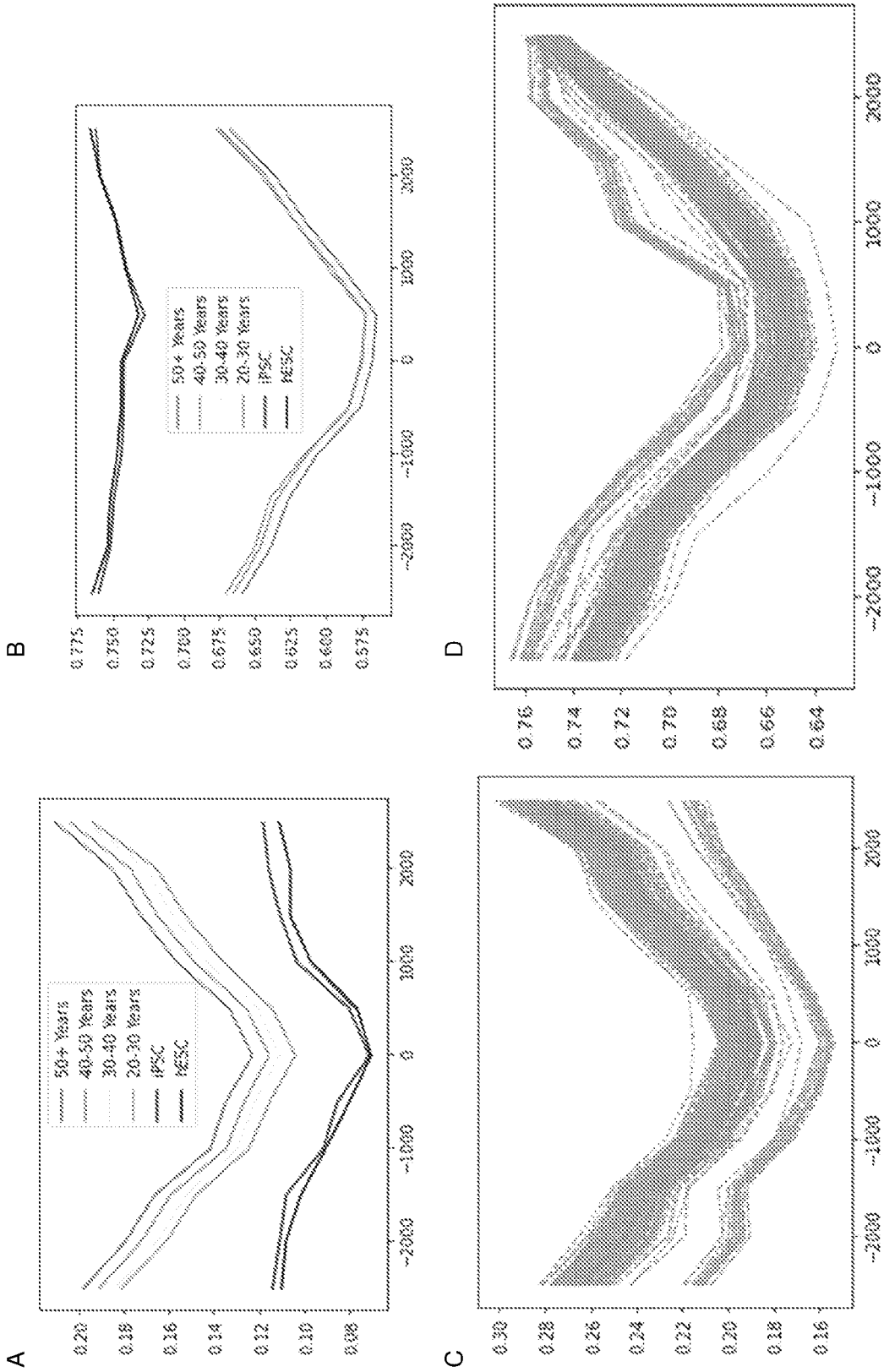


FIG. 6

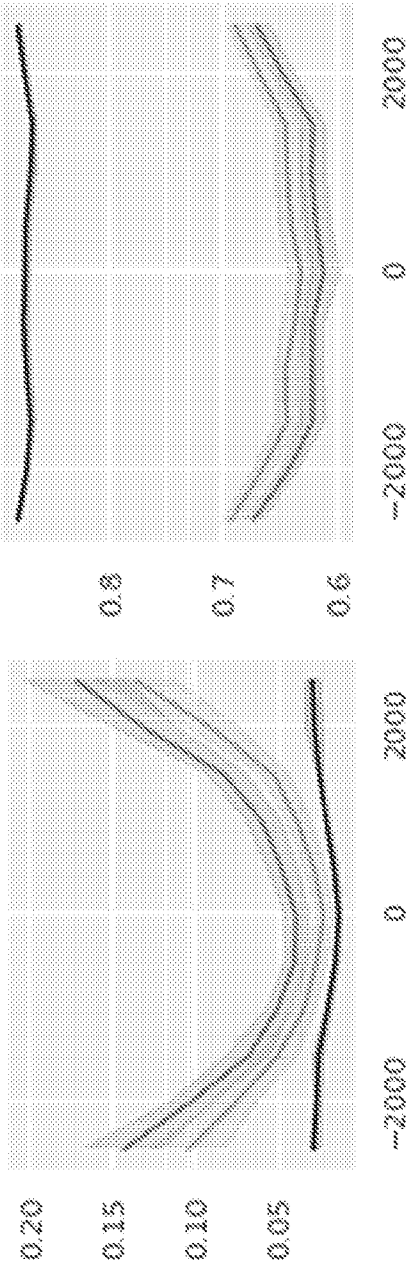


FIG. 7
Epigenetic aging index, young vs old epidermis cells

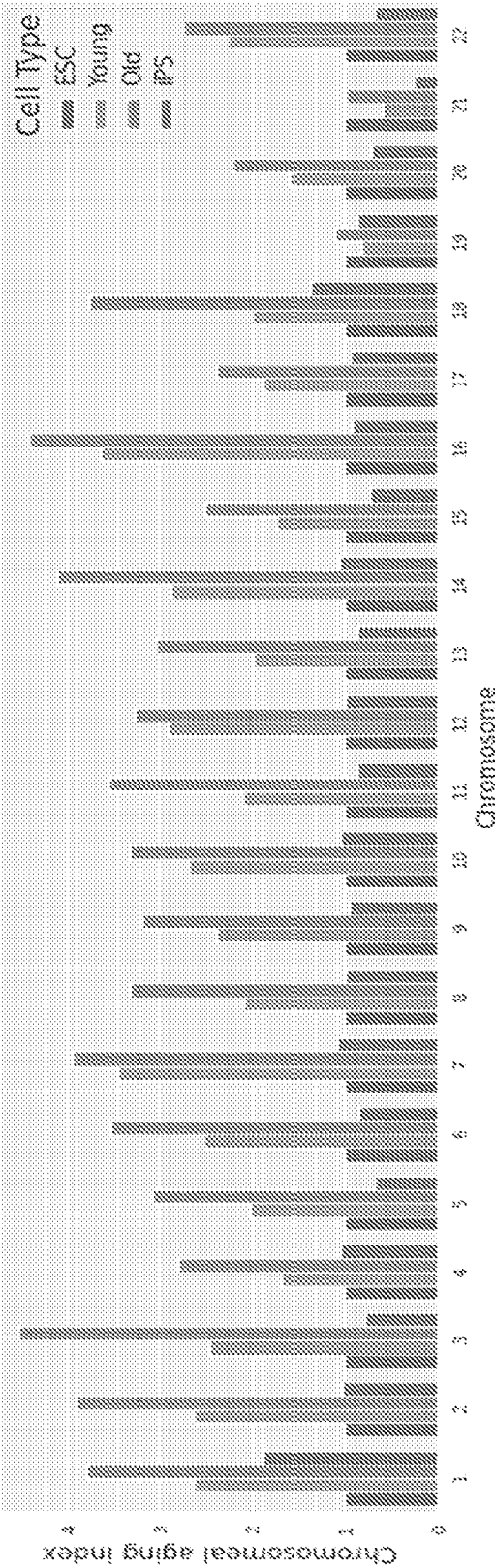
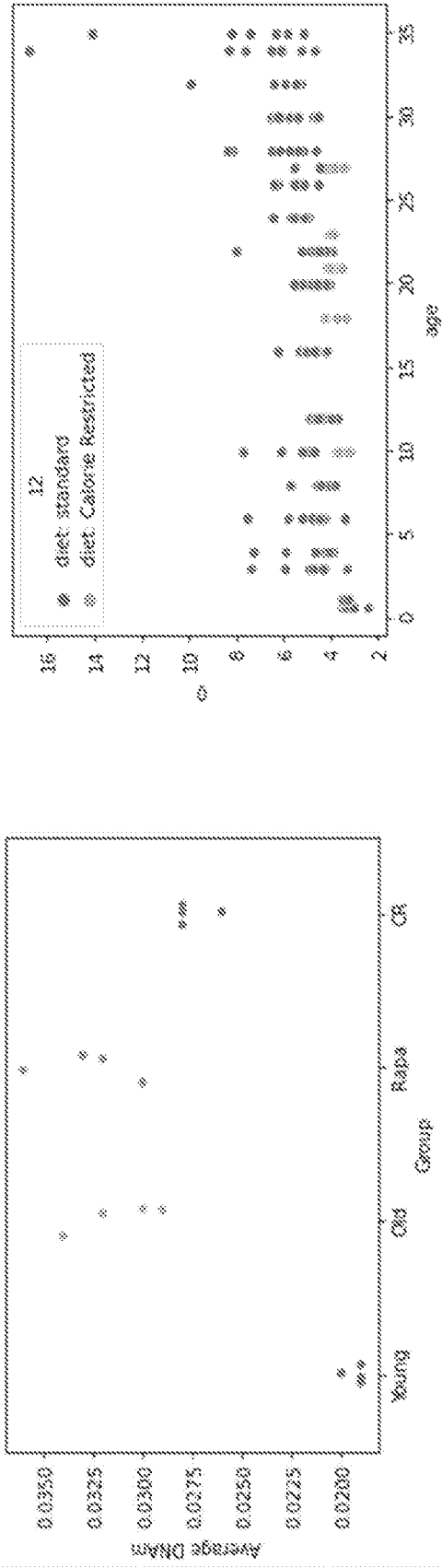
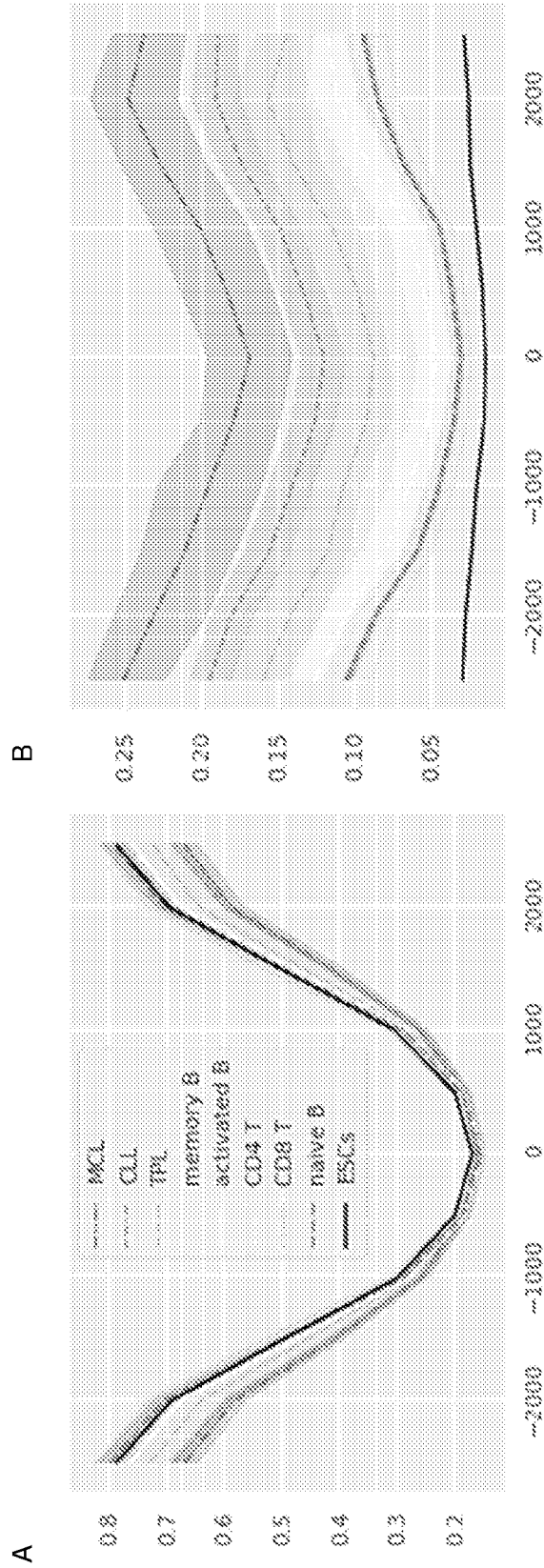


FIG. 8
AgeIndex measurement of aging intervention



AgeIndex of mice with standard vs calorie restricted diets, from 1 to 35 months

FIG. 9



INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2023/013119

A. CLASSIFICATION OF SUBJECT MATTER

IPC(8) - INV. - G16B 20/00; G16H 50/20; G16H 50/30 (2023.01)

ADD. - C12Q 1/6827; C12Q 1/6876; C12Q 1/6883 (2023.01)

CPC - INV. - G16B 20/00; G16H 50/20 (2023.02)

ADD. - C12Q 1/6827; C12Q 1/6883; C12Q 2600/136; C12Q 2600/154 (2023.02)

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)

See Search History document

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

See Search History document

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

See Search History document

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2021/0381051 A1 (THE REGENTS OF THE UNIVERSITY OF CALIFORNIA) 09 December 2021 (09.12.2021) entire document	1-5
A	US 2018/0330049 A1 (THE TRUSTEES OF COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK et al) 15 November 2018 (15.11.2018) entire document	1-5
A	DENISENKO et al. "Chromatin changes trigger laminin genes dysregulation in aging kidneys," Aging, 29 May 2018 (29.05.2018), Vol. 10, Iss. 5, Pgs. 1133-1145. entire document	1-5
A	FLORATH et al. "Cross-sectional and longitudinal changes in DNA methylation with age: an epigenome-wide analysis revealing over 60 novel age-associated CpG sites," Human Molecular Genetics, 26 October 2013 (26.10.2013), Vol. 23, Iss. 5, Pgs. 1186-1201. entire document	1-5
A	US 2020/0286625 A1 (INSILICO MEDICINE IP LIMITED) 10 September 2020 (10.09.2020) entire document	1-5
A	SCELLENBERG et al. "Replicative senescence of mesenchymal stem cells causes DNA-methylation changes which correlate with repressive histone marks," Aging, 25 September 2011 (25.09.2011), Vol. 3, Iss. 9, Pgs. 873-888. entire document	1-5
A	TRAPP et al. "Profiling epigenetic age in single cells," Nature Aging, December 2021 (12.2021), Vol. 1, Iss. 12, Pgs. 1189-1201. entire document	1-5

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

* Special categories of cited documents:

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

"D" document cited by the applicant in the international application

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

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

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

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

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

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

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

"&" document member of the same patent family

Date of the actual completion of the international search

07 April 2023

Date of mailing of the international search report

MAY 23 2023

Name and mailing address of the ISA/

Mail Stop PCT, Attn: ISA/US, Commissioner for Patents

P.O. Box 1450, Alexandria, VA 22313-1450

Facsimile No. 571-273-8300

Authorized officer

Taina Matos

Telephone No. PCT Helpdesk: 571-272-4300

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2023/013119

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.: 6-16
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:

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.:

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.