



(51) International Patent Classification:

C12Q 1/68 (2018.01) G01N 33/50 (2006.01)

(21) International Application Number:

PCT/US2018/022678

(22) International Filing Date:

15 March 2018 (15.03.2018)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

62/471,632 15 March 2017 (15.03.2017) US
62/612,188 29 December 2017 (29.12.2017) US

(71) Applicant: THE TRUSTEES OF COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK [US/US]; 412 Low Memorial Library, 535 West 116th Street, New York, NY 10027 (US).

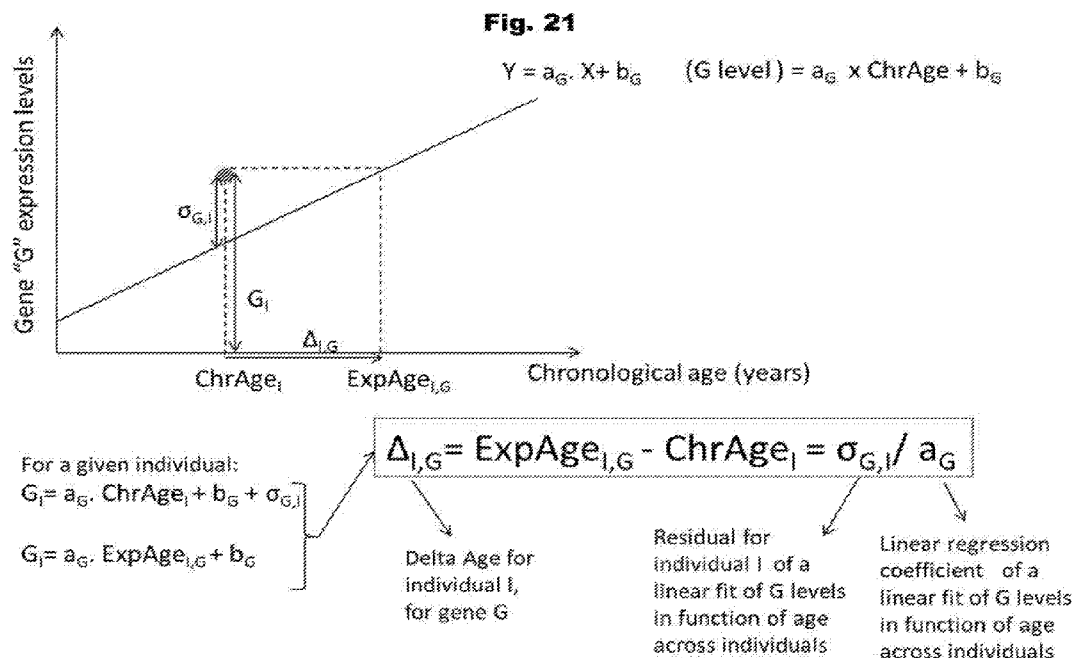
(72) Inventors: ABELIOVICH, Asa; 21 Claremont Ave. Apt. 33, New York, NY 10027 (US). RHINN, Hervé; 631 E 11th St., Apt. 24, New York, NY 10009 (US).

(74) Agent: SUPERKO, Colleen et al.; Wilmer Cutler Pickering Hale and Dorr LLP, 60 State Street, Boston, MA 02109 (US).

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

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH,

(54) Title: METHODS OF DETERMINING DIFFERENTIAL AGING AND GENETIC MODIFIERS OF GENES CORRELATED WITH A GENOTYPE OF INTEREST



(57) Abstract: The invention provides for a computer-implemented method of determining a biological age of a sample from, a subject. The invention also provides for a computer-implemented method of determining differential aging of a sample from a subject. The invention also provides for computer-implemented method of identifying one or more genetic modifiers of a phenotype associated with a genotype of interest comprising, and methods to treat phenotypes associated with genetic risk alleles.



GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

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

**METHODS OF DETERMINING DIFFERENTIAL AGING AND GENETIC
MODIFIERS OF GENES CORRELATED WITH A GENOTYPE OF INTEREST**

[0001] This application claims the benefit of and priority to U.S. Application Serial No. 62/471,632 filed March 15, 2017, and U.S. Application Serial No. 62/612188 filed December 29, 2017, the entire contents of each of which are hereby incorporated by reference in their entireties.

[0002] All patents, patent applications and publications cited herein are hereby incorporated by reference in their entirety. The disclosures of these publications in their entireties are hereby incorporated by reference into this application in order to more fully describe the state of the art as known to those skilled therein as of the date of the invention described and claimed herein.

[0003] This patent disclosure contains material that is subject to copyright protection. The copyright owner has no objection to the facsimile reproduction by anyone of the patent document or the patent disclosure as it appears in the U.S. Patent and Trademark Office patent file or records, but otherwise reserves any and all copyright rights.

BACKGROUND OF THE INVENTION

[0004] The progression of aging and its associated effects on the body is variable across the human population. While research has primarily focused on understanding extreme aging phenotypes or the identification of genetic changes associated with clear age-associated phenotypes, little progress has been made to understand the normal variability observed within the population.

SUMMARY OF THE INVENTION

[0005] The present invention provides methods of determining the biological age of a sample from a subject. The present invention also provides methods of determining the differential aging of a sample from a subject. The present invention also provides methods of determining a phenotype of a sample from a subject, wherein the phenotype is correlated with a haplotype of interest. The present invention also provides methods of determining one or more genetic modifies of a plurality of genes whose expression level is correlated with a genotype of interest. The present invention also provides methods of modifying a phenotype

associated with aging and treating, preventing, or delaying the onset of aging and cognitive decline.

[0006] In certain aspects, the invention provides a computer-implemented method of determining a biological age of a sample from a subject comprising: a) providing a gene expression level of a plurality of genes in a sample from a subject; b) determining, for each gene whose expression level is significantly correlated with chronological age, a residual value, wherein the residual value is determined by comparing the gene expression level of said gene in the sample from the subject to a linear regression of the expression level of the gene as a function of chronological age for a plurality of individuals; c) determining, for each gene whose expression level is significantly correlated with chronological age, a ratio of the residual value determined in step (b) to the linear regression coefficient for the gene; d) repeating steps (b) and (c) for each gene whose expression level is significantly correlated with chronological age; and e) integrating the ratios for each gene whose expression level is significantly correlated with chronological age, wherein the integrated ratio corresponds to the biological age of the sample from the subject.

[0007] In some embodiments, the method further comprises comparing the biological age to a chronological age of the subject to determine a differential aging value for the subject. In some embodiments, the plurality of individuals includes the subject. In some embodiments, steps (c) to (e) are performed for each individual in the plurality of individuals.

[0008] In some embodiments, the method further comprises performing a genome-wide association study (GWAS). In some embodiments, the GWAS identifies single-nucleotide polymorphism (SNP) modifiers of the differential aging value in the plurality of individuals. In some embodiments, the GWAS identifies genetic modifiers of the differential aging value in the plurality of individuals.

[0009] In some embodiments, the gene expression level is the gene expression level in the brain. In some embodiments, the gene expression level is the gene expression level in the frontal cortex. In some embodiments, the gene expression level is the gene expression level in the cerebellum.

[0010] In certain aspects, the invention provides a computer-implemented method of determining a biological age of a sample from a subject comprising: a) providing a gene expression level of a plurality of genes in a sample for a plurality of individuals; b)

performing a linear regression of the expression level of each gene of the plurality of genes as a function of chronological age of each of the plurality of individuals to determine a linear regression co-efficient for each gene of the plurality of genes; c) selecting, from the plurality of genes, the genes whose expression level is significantly correlated with chronological age; d) determining, for each gene whose expression level is significantly correlated with chronological age selected in step (c), a residual value, wherein the residual value is determined by comparing the gene expression level of said gene in the sample from the subject to the linear regression performed in step (b) for said gene; e) determining, for each gene whose expression level is significantly correlated with chronological age, a ratio of the residual value determined in step (d) to the linear regression coefficient determined in step (b); f) repeating steps (d) and (e) for each gene whose expression level is significantly correlated with chronological age selected in step (c); and g) integrating the ratios for each gene whose expression level is significantly correlated with chronological age, wherein the integrated ratio corresponds to the biological age of the sample from the subject.

[0011] In some embodiments, the method further comprises comparing the biological age to a chronological age of the subject to determine a differential aging value for the subject. In some embodiments, the plurality of individuals includes the subject. In some embodiments, steps (d) to (g) are performed for each individual in the plurality of individuals.

[0012] In some embodiments, the method further comprises performing a genome-wide association study (GWAS). In some embodiments, the GWAS identifies single-nucleotide polymorphism (SNP) modifiers of the differential aging value in the plurality of individuals. In some embodiments, the GWAS identifies genetic modifiers of the differential aging value in the plurality of individuals.

[0013] In some embodiments, the gene expression level is the gene expression level in the brain. In some embodiments, the gene expression level is the gene expression level in the frontal cortex. In some embodiments, the gene expression level is the gene expression level in the cerebellum.

[0014] In certain aspects, the invention provides a computer program product for determining a biological age of a sample from a subject comprising instructions which, when the program is executed by a computer, cause the computer to carry out the steps of: a) providing a gene expression level of a plurality of genes in a sample for a plurality of

individuals; b) determining, for each gene whose expression level is significantly correlated with chronological age, a residual value, wherein the residual value is determined by comparing the gene expression level of said gene in the sample from the subject to a linear regression of the expression level of the gene as a function of chronological age for a plurality of individuals; c) determining, for each gene whose expression level is significantly correlated with chronological age, a ratio of the residual value determined in step (b) to the linear regression coefficient for the gene; d) repeating steps (b) and (c) for each gene whose expression level is significantly correlated with chronological age; and e) integrating the ratios for each gene whose expression level is significantly correlated with chronological age, wherein the integrated ratio corresponds to the biological age of the sample from the subject.

[0015] In some embodiments, the computer program product further comprises carrying out the step of: comparing the biological age to a chronological age of the subject to determine a differential aging value for the subject. In some embodiments, the plurality of individuals includes the subject. In some embodiments, steps (c) to (e) are performed for each individual in the plurality of individuals.

[0016] In some embodiments, the computer program product further comprises carrying out the step of performing a genome-wide association study (GWAS). In some embodiments, the GWAS identifies single-nucleotide polymorphism (SNP) modifiers of the differential aging value in the plurality of individuals. In some embodiments, the GWAS identifies genetic modifiers of the differential aging value in the plurality of individuals.

[0017] In some embodiments, the gene expression level is the gene expression level in the brain. In some embodiments, the gene expression level is the gene expression level in the frontal cortex. In some embodiments, the gene expression level is the gene expression level in the cerebellum.

[0018] In certain aspects, the invention provides a computer program product for determining a biological age of a sample from a subject comprising instructions which, when the program is executed by a computer, cause the computer to carry out the steps of: a) providing a gene expression level of a plurality of genes in a sample for a plurality of individuals; b) performing a linear regression of the expression level of each gene of the plurality of genes as a function of chronological age of each of the plurality of individuals to determine a linear regression co-efficient for each gene of the plurality of genes; c) selecting,

from the plurality of genes, the genes whose expression level is significantly correlated with chronological age; d) determining, for each gene whose expression level is significantly correlated with chronological age selected in step (c), a residual value, wherein the residual value is determined by comparing the gene expression level of said gene in the sample from the subject to the linear regression performed in step (b) for said gene; e) determining, for each gene whose expression level is significantly correlated with chronological age, a ratio of the residual value determined in step (d) to the linear regression coefficient determined in step (b); f) repeating steps (d) and (e) for each gene whose expression level is significantly correlated with chronological age selected in step (c); and g) integrating the ratios for each gene whose expression level is significantly correlated with chronological age, wherein the integrated ratio corresponds to the biological age of the sample from the subject.

[0019] In some embodiments, the computer program product further comprises carrying out the step of: comparing the biological age to a chronological age of the subject to determine a differential aging value for the subject. In some embodiments, the plurality of individuals includes the subject. In some embodiments, steps (d) to (g) are performed for each individual in the plurality of individuals.

[0020] In some embodiments, the computer program product further comprises carrying out the step of performing a genome-wide association study (GWAS). In some embodiments, the GWAS identifies single-nucleotide polymorphism (SNP) modifiers of the differential aging value in the plurality of individuals. In some embodiments, the GWAS identifies genetic modifiers of the differential aging value in the plurality of individuals.

[0021] In some embodiments, the gene expression level is the gene expression level in the brain. In some embodiments, the gene expression level is the gene expression level in the frontal cortex. In some embodiments, the gene expression level is the gene expression level in the cerebellum.

[0022] In certain aspects, the invention provides a computer-implemented method of determining a phenotype of a sample from a subject, wherein the phenotype is correlated with a haplotype of interest, the method comprising: a) providing a gene expression level of a plurality of genes in a sample for a plurality of individuals, wherein each individual's haplotype of interest is known; b) performing a linear regression of the expression level of each gene of the plurality of genes as a function of the haplotype of interest for each of the

plurality of individuals to determine a linear regression co-efficient for each gene of the plurality of genes; c) selecting, from the plurality of genes, the genes whose expression level is significantly correlated with the haplotype of interest; d) determining, for each gene whose expression level is significantly correlated with the haplotype of interest selected in step (c), a residual value, wherein the residual value is determined by comparing the gene expression level of said gene in a sample from a subject to the linear regression performed in step (b) for said gene; e) determining, for each gene whose expression level is significantly correlated with the haplotype of interest, a ratio of the residual value determined in step (d) to the linear regression coefficient determined in step (b); f) repeating steps (d) and (e) for each gene whose expression level is significantly correlated with the haplotype of interest selected in step (c); and g) integrating the ratios for each gene whose expression level is significantly correlated with the haplotype of interest, wherein the integrated ratio corresponds to the phenotype of the sample from the subject.

[0023] In some embodiments, the plurality of individuals includes the subject. In some embodiments, steps (d) to (g) are performed for each individual in the plurality of individuals.

[0024] In some embodiments, the method further comprises: h) performing a genome-wide association study (GWAS). In some embodiments, the GWAS identifies single-nucleotide polymorphism (SNP) modifiers of the phenotype correlated with the haplotype of interest in the plurality of individuals. In some embodiments, the GWAS identifies genetic modifiers of the phenotype in the plurality of individuals. In some embodiments, the haplotype of interest is defined as 0, 1, or 2 allele copies. In some embodiments, the allele copies are determined by SNP genotyping. In some embodiments, the phenotype correlated with the haplotype of interest is an expression level of a plurality of genes.

[0025] In some embodiments, the gene expression level is the gene expression level in the brain. In some embodiments, the gene expression level is the gene expression level in the frontal cortex. In some embodiments, the gene expression level is the gene expression level in the cerebellum.

[0026] In certain aspects, the invention provides a computer program product for determining the phenotype of a sample from a subject, wherein the phenotype is correlated with a haplotype of interest, comprising instructions which, when the program is executed by a computer, cause the computer to carry out the steps of: a) providing a gene expression level

of a plurality of genes in a sample for a plurality of individuals, wherein each individual's haplotype of interest is known; b) performing a linear regression of the expression level of each gene of the plurality of genes as a function of the haplotype of interest for each of the plurality of individuals to determine a linear regression co-efficient for each gene of the plurality of genes; c) selecting, from the plurality of genes, the genes whose expression level is significantly correlated with the haplotype of interest; d) determining, for each gene whose expression level is significantly correlated with the haplotype of interest selected in step (c), a residual value, wherein the residual value is determined by comparing the gene expression level of said gene in a sample from a subject to the linear regression performed in step (b) for said gene; e) determining, for each gene whose expression level is significantly correlated with the haplotype of interest, a ratio of the residual value determined in step (d) to the linear regression coefficient determined in step (b); f) repeating steps (d) and (e) for each gene whose expression level is significantly correlated with the haplotype of interest selected in step (c); and g) integrating the ratios for each gene whose expression level is significantly correlated with the haplotype of interest, wherein the integrated ratio corresponds to the phenotype of the sample from the subject.

[0027] In some embodiments, the plurality of individuals includes the subject. In some embodiments, wherein steps (d) to (g) are performed for each individual in the plurality of individuals.

[0028] In some embodiments, the computer program product further comprises carrying out the step of: h) performing a genome-wide association study (GWAS). In some embodiments, the GWAS identifies single-nucleotide polymorphism (SNP) modifiers of the phenotype in the plurality of individuals. In some embodiments, the GWAS identifies genetic modifiers of the phenotype in the plurality of individuals. In some embodiments, the haplotype of interest is defined as 0, 1, or 2 allele copies. In some embodiments, the allele copies are determined by SNP genotyping. In some embodiments, the phenotype correlated with a haplotype of interest is an expression level of a plurality of genes.

[0029] In some embodiments, the gene expression level is the gene expression level in the brain. In some embodiments, the gene expression level is the gene expression level in the frontal cortex. In some embodiments, the gene expression level is the gene expression level in the cerebellum.

[0030] In certain aspects, the invention provides a computer-implemented method of identifying one or more genetic modifiers of a plurality of genes whose expression level is correlated with a genotype of interest comprising: a) providing a gene expression level of a plurality of genes in a sample for a plurality of individuals, wherein each individual's genotype of interest is known; b) performing a linear regression of the expression level of each gene of the plurality of genes as a function of the genotype of interest of each of the plurality of individuals to determine a linear regression co-efficient for each gene of the plurality of genes; c) selecting, from the plurality of genes, the genes whose expression level is significantly correlated with the genotype of interest; d) determining, for each gene whose expression level is significantly correlated with the genotype of interest selected in step (c), a residual value, wherein the residual value is determined by comparing the gene expression level of said gene in a sample from a subject to the linear regression performed in step (b) for said gene; e) determining, for each gene whose expression level is significantly correlated with the genotype of interest, a ratio of the residual value determined in step (d) to the linear regression coefficient determined in step (b); f) repeating steps (d) and (e) for each gene whose expression level is significantly correlated with the genotype of interest selected in step (c); g) integrating the ratios for each gene significantly correlated with the genotype of interest; h) performing steps (d) to (g) for each individual in the plurality of individuals; and j) identifying genetic modifiers that modulate the integrated ratio value determined in step (g).

[0031] In some embodiments, the genetic modifiers that modulate the integrated ratio value determined in step (g) are determined by performing a genome-wide association study (GWAS). In some embodiments, the plurality of individuals includes the subject.

[0032] In some embodiments, the GWAS identifies single-nucleotide polymorphism (SNP) modifiers of the integrated ratio value determined in step (g) in the plurality of individuals. In some embodiments, the genetic modifier is a single-nucleotide polymorphism (SNP). In some embodiments, the genotype of interest is the risk allele of TMEM106B. In some embodiments, the risk allele of TMEM106B is and A at SNP rs1990622. In some embodiments, the genotype of interest is a risk allele associated with a disease or disorder. In some embodiments, the genotype of interest is a non-risk allele. In some embodiments, the genotype of interest is a haplotype.

[0033] In some embodiments, the gene expression level is the gene expression level in the brain. In some embodiments, the gene expression level is the gene expression level in the frontal cortex. In some embodiments, the gene expression level is the gene expression level in the cerebellum.

[0034] In some embodiments, the plurality of individuals are healthy individuals. In some embodiments, the plurality of individuals have a disease or disorder. In some embodiments, the plurality of individuals have a neurodegenerative disease. In some embodiments, the neurodegenerative disease is Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, or Fronto-temporal dementia.

[0035] In certain aspects, the invention provides a computer program product for identifying one or more genetic modifiers of a plurality of genes whose expression level is correlated with a genotype of interest comprising instructions which, when the program is executed by a computer, cause the computer to carry out the steps of: a) providing a gene expression level of a plurality of genes in a sample for a plurality of individuals, wherein each individual's genotype of interest is known; b) performing a linear regression of the expression level of each gene of the plurality of genes as a function of the genotype of interest of each of the plurality of individuals to determine a linear regression co-efficient for each gene of the plurality of genes; c) selecting, from the plurality of genes, the genes whose expression level is significantly correlated with the genotype of interest; d) determining, for each gene whose expression level is significantly correlated with the genotype of interest selected in step (c), a residual value, wherein the residual value is determined by comparing the gene expression level of said gene in a sample from a subject to the linear regression performed in step (b) for said gene; e) determining, for each gene whose expression level is significantly correlated with the genotype of interest, a ratio of the residual value determined in step (d) to the linear regression coefficient determined in step (b); f) repeating steps (d) and (e) for each gene whose expression level is significantly correlated with the genotype of interest selected in step (c); g) integrating the ratios for each gene significantly correlated with the genotype of interest; h) performing steps (d) to (g) for each individual in the plurality of individuals; and j) identifying genetic modifiers that modulate the integrated ratio value determined in step (g).

[0036] In some embodiments, the genetic modifiers that modulate the integrated ratio value determined in step (g) are determined by performing a genome-wide association study

(GWAS). In some embodiments, the plurality of individuals includes the subject. In some embodiments, the GWAS identifies single-nucleotide polymorphism (SNP) modifiers of the integrated ratio value determined in step (g) in the plurality of individuals. In some embodiments, the genetic modifier is a single-nucleotide polymorphism (SNP). In some embodiments, the genotype of interest is the risk allele of TMEM106B. In some embodiments, the risk allele of TMEM106B is and A at SNP rs1990622. In some embodiments, the genotype of interest is a risk allele associated with a disease or disorder. In some embodiments, the genotype of interest is a non-risk allele. In some embodiments, the genotype of interest is a haplotype.

[0037] In some embodiments, the gene expression level is the gene expression level in the brain. In some embodiments, the gene expression level is the gene expression level in the frontal cortex. In some embodiments, the gene expression level is the gene expression level in the cerebellum.

[0038] In some embodiments, the plurality of individuals are healthy individuals. In some embodiments, the plurality of individuals have a disease or disorder. In some embodiments, the plurality of individuals have a neurodegenerative disease. In some embodiments, the neurodegenerative disease is Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, or Fronto-temporal dementia.

[0039] In certain aspects, the invention provides a method of modifying a phenotype associated with a TMEM106B risk allele in a subject in need thereof, the method comprising administering an effective amount of an IL2 modulator, and IL2RA modulator, an IL2RB modulator, an IL2RG modulator, an IL15 modulator, an IL15RA modulator, or a combination thereof to the subject.

[0040] In some embodiments, the subject is administered an IL2 modulator. In some embodiments, the subject is administered an IL2RA modulator. In some embodiments, the subject is administered an IL2RB modulator. In some embodiments, the subject is administered an IL2RG modulator. In some embodiments, the subject is administered an IL15 modulator. In some embodiments, the subject is administered an IL15RA modulator. In some embodiments, the subject is homozygous for the TMEM106B risk allele. In some embodiments, the subject is heterozygous for the TMEM106B risk allele. In some embodiments, the TMEM106B risk allele is an A at SNP rs1990622. In some embodiments,

the subject is homozygous for the TMEM106B protective allele. In some embodiments, the TMEM106B protective allele is a G at SNP rs1990622. In some embodiments, the IL2RA modulator increases expression of a IL2RA protective allele, or decreases expression of a IL2RA risk allele, or a combination thereof. In some embodiments, the IL2RA protective allele is an A at SNP rs12722515. In some embodiments, the IL2RA risk allele is an C at SNP rs12722515.

[0041] In certain aspects, the invention provides a method of treating, preventing, or delaying the onset of aging in a subject in need thereof, the method comprising administering an effective amount of an IL2 modulator, and IL2RA modulator, an IL2RB modulator, an IL2RG modulator, an IL15 modulator, an IL15RA modulator, a TMEM106B modulator, or a combination thereof to the subject.

[0042] In certain aspects, the invention provides a method of treating, preventing, or delaying the onset of cognitive decline in a subject in need thereof, the method comprising administering an effective amount of an IL2 modulator, and IL2RA modulator, an IL2RB modulator, an IL2RG modulator, an IL15 modulator, an IL15RA modulator, a TMEM106B modulator, or a combination thereof to the subject.

[0043] In some embodiments, the modulation increases expression of a TMEM106B protective allele. In some embodiments, the modulation decreases the expression of the TMEM106B risk allele. In some embodiments, the TMEM106B risk allele is an A at SNP rs1990622. In some embodiments, the TMEM106B protective allele is a G at SNP rs1990622.

[0044] In some embodiments, the phenotype associated with a TMEM106B risk allele is a plurality of genes, and their expression levels, associated with the TMEM106B risk allele. In some embodiments, the phenotype associated with a TMEM106B risk allele is reduced and a phenotype associated with a TMEM106B protective allele is increased. In some embodiments, the phenotype associated with a TMEM106B risk allele is a plurality of genes whose expression level is correlated with the TMEM106B risk allele.

[0045] As would be apparent to one of ordinary skill in the art, any method or composition described herein can be implemented with respect to any other method or composition described herein.

[0046] These, and other, embodiments of the invention will be better appreciated and understood when considered in conjunction with the following description and the accompanying drawings. It should be understood, however, that the following description, while indicating various embodiments of the invention and numerous specific details thereof, is given by way of illustration and not of limitation. Many substitutions, modifications, additions and/or rearrangements may be made within the scope of the invention without departing from the spirit thereof, and the invention includes all such substitutions, modifications, additions and/or rearrangements.

BRIEF DESCRIPTION OF THE FIGURES

[0047] To conform to the requirements for PCT patent applications, many of the figures presented herein are black and white representations of images originally created in color.

[0048] **Figs. 1A-B.** Aging rates are heterogeneous across individuals within a cohort. (A) Schematic representation of variability in the relative rate and pattern of progression of age-associated traits as a function of time. In this hypothetical example, a generic aging trait progresses more rapidly in individual 2 than individual 1, whereas individual 3 displays a bimodal pattern. (B) Schematics representation of Δ -aging analysis. Each dot represents, for a single individual, the tissue expression level (x-axis) of a hypothetical age-dependent gene as a function of chronological age (y-axis). In this example, expression levels are positively correlated with chronological age across the cohort, as shown by the regression line. Individuals that display an expression level higher than predicted for their chronological age, such as the sample highlighted in red, exhibit an estimated biological age higher than their chronological age (positive Δ -aging). In contrast, samples that display an expression level lower than predicted for their chronological age, such as the sample highlighted in blue, would be associated with an estimated biological age lower than the chronological age (negative Δ -aging). Integration across all age-associated genes constitutes the aggregate Δ -aging for that individual.

[0049] **Figs. 2A-C.** Transcriptome-wide analysis identifies age-associated gene expression changes in human brain. (A) Age-associated changes in the aggregated expression levels of gene sets that typify different human CNS cell types in prefrontal cortex, as labeled. Transcriptome-wide meta-analysis of age-associated gene expression levels was conducted using 4 prefrontal cortex datasets; n=716 neurodegenerative-disease free individuals > 25yo. Mean age-associated fold-changes across gene sets members are plotted (n=21 per group).

Error bars are SEM. *: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$ for linear association between age and aggregated expression levels in a meta-analysis across the 4 expression datasets. Neurons (red): $Z = -3.93$, $p = 8.47 \times 10^{-5}$; Neurons/OPCs (burgundy): $Z = -4.24$, $p = 2.24 \times 10^{-5}$; OPCs (purple) $Z = -2.27$, $p = 2.33 \times 10^{-2}$; Neurons/Astrocytes (orange): $Z = -1.89$, $p = 5.81 \times 10^{-2}$; Endothelial cells (pink): $Z = 4.02$, $p = 5.78 \times 10^{-5}$; Microglia (yellow): $Z = 2.93$, $p = 3.39 \times 10^{-3}$; Oligodendrocytes (green): $Z = 3.97$, $p = 7.31 \times 10^{-5}$; Astrocytes (blue): $Z = 5.29$, $p = 1.25 \times 10^{-7}$. (B) (C) Expression levels of PNOC, a representative neuronal gene that is decreased in expression with aging (B, red) and GFAP, a representative astrocyte gene that is increased in expression with aging (C, blue) as a function of age. Each dot represents an individual sample; $n = 128$ individuals of age > 25 yo from the Braincloud dataset; regression lines are shown.

[0050] Figs. 3A-F. Genome-wide association study identifies a genetic determinant of aging rate in human frontal cortex at the TMEM106B locus. (A) Schematic of the genetic analysis of modifiers of Δ -aging in human frontal cortex. (B) Tabular presentation of the associations observed between rs1990622 genotype at TMEM106B and Δ -aging through stages of the GWAS. Effects are expressed in terms of years per minor allele load. See Example 1 Methods for details on the statistical analyses. (C) Manhattan plot representing the association between Δ -aging, as quantified in frontal cortex tissue samples from older adults, and each of 468,129 common SNP variants (meta-analysis of 5 cohorts; Discovery+Replication as in Fig. 3B and Fig. 15; $n = 910$). The red line corresponds to a threshold ($p < 1.06 \times 10^{-7}$) for genome wide significance after Bonferroni correction for the multiple SNPs tested. Highlighted in red are the SNPs in the region of interest (chr7:11,783,787-12,783,787) surrounding the association peaks at the TMEM106B (rs1990622) and GRN loci. (D) Local Manhattan plot in the region of interest of the TMEM106B gene locus, representing the genome-wide association p-value between common genetic variants and Δ -aging in older adults' frontal cortex samples in a meta-analysis of 5 cohorts (Discovery+Replication, $n = 910$) after local imputation. (E) Effect of rs1990622 allele load on Δ -aging value in a meta-analysis of 910 tissue samples from older neurodegenerative-free individuals. Homozygosity for the minor protective (PP) allele is associated with a 12-year decrease in Δ -aging, relative to homozygosity of the risk (RR) allele. $N = 182$, 438 and 290 for the PP, PR and RR genotypes. Mean values are presented. Error bars are SEM. +++: $p = 2.74 \times 10^{-22}$ for the effect of "R" allelic load on Δ -aging by Kruskal-Wallis test, chi-square 99.30 with 1 degree of freedom. (F) Schematic representation of the effect of TMEM106B

rs1990622 genotype on brain aging trajectories. Risk allele carriers display an accelerated aging phenotype in late life (>65 yo).

[0051] Figs. 4A-C. TMEM106B and age affect cognitive function in elderly individuals. (A) (B) Carriers of the TMEM106B Δ -aging risk-associated haplotype display age-associated cognitive deficits, as assessed by Mini Mental State Examination in a cross-sectional cohort (Long Life Study) of individuals. Consistent with the Δ -aging findings, the deficits are seen in older (65 to 80yo) but not younger (50 to 65yo) individuals. rs1060700 was used a proxy for rs1990622 (pairwise linkage disequilibrium: $R^2=1$, $D'=1$ in the CEU 1000 Genomes datasets). $N=262$, 744 and 541, $N=192,429$ and 321 for the homozygous protective allele (PP), heterozygous (PR) and homozygous risk (RR) genotypes in the 50-65yo and 65-80yo age groups, respectively. Mean values are presented. Error bars are SEM. ++: $p<0.01$ by linear regression for the additive allelic load, correcting for age, gender and 3 principal eigenvectors for population stratification. (C) Carriers of the TMEM106B Δ -aging risk-associated haplotype display age-associated cognitive deficits, as assessed by a memory recall score in the Health and Retirement Study. The Differential-aging associated genotype was queried using rs1060700 as a proxy for rs1990622 (pairwise linkage disequilibrium: $R^2=1$, $D'=1$ in the CEU 1000 Genomes datasets).

[0052] Figs. 5A-C. The effect of TMEM106B genetic variant on the transcriptome appears similar to the effect of age exclusively in elderly individuals. (A) Venn Diagram representing the number of genes whose levels correlate with age or TMEM106B risk allele load in cohorts of older ($n=413$) or younger ($n=303$) adults from the “Discovery” datasets (with an association p -value <0.01 for each factor with same direction of changes, Fig. 3B, Fig. 15, and meta-analysis). The effect of the TMEM106B risk allele appears very different in younger versus older individuals: its impact on the transcriptome is potentiated in the individuals over 65 years old, in which its global signature resembles the one associated to chronological age. By contrast, the impact of chronological age appears marked in both individuals below or above 65 yo. (B)(C). Dot plots display, for each of 15,139 individual genes (each represented by a single dot), the degree to which gene expression is correlated with TMEM106B risk allele load (y-axis) versus the degree to which gene expression is correlated with chronological age (x-axis). Separate analyses are presented of older (b; $n=413$) or younger (c; $n=303$) adults from the “Discovery” datasets (see Figs. 3B and 15). Z-scores represent the statistical significance and direction of the age- or genotype-associated correlations. Z-values of 1.96, 2.56 and 3.29 correspond to p -values of 0.05, 0.01 and 0.001 respectively.

Regression lines (in red) show that in older individuals (C), those genes that are more highly correlated (positively or negatively) in expression with TMEM106b risk allele load (y-axis) are also more highly correlated with age (x-axis). Individual genes characteristic of neurons (red), astrocytes (blue) and microglia (yellow) are highlighted as examples.

[0053] Figs. 6A-D. TMEM106B modulates innate immune cell inflammatory polarization.

(A) Schematic of the inflammatory polarization of myeloid cells by pro- or anti-inflammatory factors. (B) Aggregate cerebral cortical expression levels of gene sets associated with M1-type (red, n=39 genes) or M2-type (green n=39 genes) microglia were quantified for neurodegenerative-disease free individuals homozygous for the risk (RR) or protective (PP) alleles, or heterozygous (PR) for these alleles, at the TMEM106B locus. Expression was found to increase steadily in all groups over time (presented as percent change per decade); for the M2 gene set, the increase was significantly less rapid in homozygous carriers of the risk allele, relative carriers of the protective allele. For each group, median values across the datasets are presented, (n=4 datasets from the Discovery stage, with 716 individuals in total; see Fig. 15). Error bars are SEM. †††: $p < 0.001$, n.s: $p > 0.05$ for the effect of rs1990622 risk allele load by ANOVA, correcting for dataset-specific effects ($p = 9.41 \times 10^{-4}$, $F = 21.41$ for 1 degree of freedom and $p = 0.737$, $F = 0.119$ for 1 degree of freedom for M2 and M1 respectively). (C) Effect of rs1990622 genotype or LPS treatment on the aggregate expression levels of M1-selective and M2-selective genes in human monocyte-derived dendritic cells, as quantified by Nanostring (GEO GSE53165). Data are presented as a ratio of the M1/M2 aggregate levels; higher values are thus associated with a pro-inflammatory state. Mean values are presented. Error bars are SEM. N=29/82/49 and 13/51/23 for the RR/PR/PP genotypes in the Unstimulated and LPS-treated groups, respectively. ***: $p < 0.001$ treatment effect, +: $p < 0.05$ allelic load association by Kruskal-Wallis test ($p = 1.08 \times 10^{-32}$, chi-square= 141.78 for 1 degree of freedom and $p = 1.09 \times 10^{-2}$, chi-square= 9.02 for 1 degree of freedom respectively). (D) Effects of rs1990622 genotype or LPS treatment on TMEM106B expression levels in human monocyte-derived dendritic cells, as measured by Affymetrix Human Gene 1.0 ST Array (GSE53166). Mean values are presented. Error bars are SEM. N=2/14/7 and 2/14/8 for the RR/PR/PP genotypes in the Unstimulated and LPS-treated groups respectively. ***: $p < 0.001$ for treatment effect by ANOVA, ++: $p < 0.01$ for allelic load association in unstimulated by ANOVA ($p = 5.0 \times 10^{-3}$, $F = 9.85$ for 1 degree of freedom and $p = 1.13 \times 10^{-12}$, $F = 96.624$ for 1 degree of freedom).

[0054] Figs. 7A-B. Non-genetic factors can modify Δ -aging. (A) Δ -aging values observed in frontal cortex or cerebellum tissue samples derived from neurological disease-free, AD, or

HD cohorts (Harvard Brain Bank). Δ -aging values are presented relative to the values observed in unaffected individuals for each brain region. Mean values are presented. Error bars are SEM. N=154,345,170,124,269 and 140 ***: $p<0.001$, *: $p<0.05$ by Kruskal-Wallis test followed by Dunn's multiple comparisons test with Und group in the same tissue; †††: $p<0.001$ by Kruskal-Wallis followed by Dunn's multiple comparisons test for comparison with Frontal cortex tissue in the same disease category. ($p=3.91\times 10^{-18}$ chi-square= 91.12 with 5 degrees of freedom for Kruskal-Wallis, with Dunn's test p-values of $p=3.72\times 10^{-13}$ and 9.36×10^{-8} for AD vs Und and HD vs Und respectively in Frontal Cortex, $p=1.74\times 10^{-2}$ for AD vs Und in Cerebellum and $p=2.71\times 10^{-7}$ and 1.14×10^{-4} for Cerebellum vs Frontal Cortex in AD and HD respectively). (B) Longitudinal analysis of Δ -aging in serial muscle tissue biopsies from elderly individuals before and after a 6mo-long vigorous exercise routine program (GEO GSE8479). Black dots represent muscle samples from individuals that did not perform an exercise routine; these data were used to define the pattern of age-associated gene expression changes within the cohort. Red dots represent muscle samples taken from elderly individuals before the exercise regimen, whereas blue dots represent muscle samples from the same individuals after training. The mean Δ -aging shift observed after exercise training is -24.43 ± 1.94 years (sem, $n=14$ individuals, $p=1.14\times 10^{-8}$ by paired two-tailed t-test).

[0055] **Figs. 8A-B.** Related to Figs. 1A-B. (A) Determination of delta-age for a given gene. The delta-age for individual I related to gene G (Δ_{IG}) is determined as the ratio between the residual value ($\sigma_{G,I}$) for individual I in a linear fit of G levels in function of age across individuals onto the age axis, and the linear regression coefficient (a_G) of such fit. (B) The global delta age for a given individual is evaluated by aggregating its delta values related individuals' genes across all the age-associated genes.

[0056] **Fig. 9.** Related to Figs. 3A-F. Effect of rs1990622 allele load on Delta-ageing value in individual datasets as indicated, in individuals aged of more or less than 65yo (upper and lower panels respectively). Mean values are presented. Error bars are SEM. N as indicated on individual panels.

[0057] **Fig. 10.** Related to Figs. 3A-F. Locus zoom Manhattan plot representing the genome-wide association p-value between common genetic variants and Delta-Age in older adults' frontal cortex samples in a meta-analysis of 5 cohorts (Discovery+Replication, $n=910$) after local imputation at the GRN locus, highlighted in blue on the genome-wide analysis presented in Fig. 3C.

[0058] **Fig. 11.** Related to Figs. 3A-F. Dot-plot of apparent biological age as calculated by the Delta-aging procedure in function of actual chronological age in individuals from the

Discovery cohorts aged of more than 65 yo (n=413). Individuals are colored in function of their TMEM106B rs1990622 genotypes (green for the rs1990622 protective allele homozygotes, red for the rs1990622 risk allele homozygotes, grey for the heterozygotes). The effect of TMEM106B rs1990622 genotype on brain aging trajectories in individuals is highlighted by the linear regression lines (dashed lines with corresponding colors for each of the 3 genotypes).

[0059] **Figs. 12A-B.** Related to Figs. 5A-C. (A). Heat map representing the correlation and associated p-value between age, TMEM106B rs1990622 risk allele load (RAL) or gender with levels of genes from 5 clusters, labelled yellow, blue, green, brown and turquoise, identified in a hypothesis-free fashion by WGCNA in a gene expression array dataset of 187 cortical brain samples from neurodegenerative disease-free older individuals, and found to be enriched for genes associated with either microglia, astrocytes, oligodendrocytes, neuron gene, with respective enrichment p-values of $p=3.4 \times 10^{-33}$, 2.85×10^{-14} , 5.34×10^{-30} , 6.6×10^{-4} and $p=1.2 \times 10^{-2}$ (See Example 1 Methods). rs1990622 risk allele load, or increased age, was associated with an upregulation of aggregated microglial gene cluster expression. Increased age was also associated with an overall reduction in aggregated gene expression in the neuronal cluster, whereas gender was not associated with an alteration in the expression of any gene cluster. (B). Radar plot displaying the correlation between the aggregate expression levels of gene sets associated with different CNS cell types, as labeled, and either chronological age or rs1990622 risk-allele load. Z-values (presented along the spokes, with high correlation on the perimeter and high anti-correlation at center) for the linear correlations with age in younger (in orange; n=303) or older adults (in green; n=413), or with rs1990622 risk allele load in younger (blue; n=303) or older (purple; n=413) adults is shown.

[0060] **Figs. 13A-D.** Related to Figs. 6A-D. (A) Heatmap presenting the effect on gene expression of rs1990622 genotype in human dendritic cells, or of M1-/M2- polarization in human macrophages for genes that were found to be significantly affected by either M1- or M2- polarization and present in among the genes tested in human dendritic cells with the Nanostring assay. (B)(C)(D): Effect of rs1990622 genotype and different inflammatory treatment on the combined expression levels of M1-macrophage genes (C) M2- (D) or over M2 over M1 (B) macrophage genes in human monocyte-derived dendritic cells from individual European ascent measured by Nanostring (GEO GSE53165). The effect of higher values is associated to a proinflammatory effect. Mean values are presented. Error bars are SEM. N=29,82,49; 13,51, 23; 28,71,50; 25,73,45; 5,73, 45.

[0061] **Figs. 14A-F.** Related to Figs. 7A-B. Genome-wide Manhattan plots representing the genome-wide association p-value between common genetic variants and Delta-Age in meta-analysis of 4 cohorts (BrainEqtl, HBTRC Discovery, HBTRC AD and HBTRC HD, n = 924) for which samples from both brain cortex and cerebellum were available from the same donors. Association analysis were carried for either Delta-age values as defined in the frontal cortex on the basis of the expression values observed in those samples (A-C), or in the cerebellum (d-e). Such analysis was carried either in all adults (A, B) or specifically among older or younger individuals (C, D and E, F respectively). The red line corresponds to a threshold ($p < 1.06 \times 10^{-7}$) for genome wide significance after Bonferroni correction for the 468,129 SNPs tested. Highlighted in green are the SNPs in the region of interest (chr7:11,783,787-12,783,787) surrounding the association peak at rs1990622.

[0062] **Fig. 15.** Associations observed between rs1990622 genotype at TMEM106B and Differential-aging in the datasets. Effect are expressed in years per minor allele load.

[0063] **Fig. 16.** Individuals, represented as dots, are plotted as a function of their chronological age (X-axis) and their measured expression for gene G (Y-axis). The dotted line corresponds to the regression line for Gene G expression levels as a function of chronological age across the entire cohort.

[0064] **Fig. 17.** Linear regression across individuals of the expression level of a gene G in function of chronological age yields a regression line. For a given individual the expression level of gene G is shown.

[0065] **Fig. 18.** For a given individual the global delta-age is obtained by integration of all the gene-specific Delta-Age over all genes whose expression levels are found to be correlated with chronological age during the original linear regression.

[0066] **Fig. 19.** Aging rate as a differential in an age-related trait. In red: individuals with a level higher than one would expect for their age: "looking older." In blue: individuals with a level lower than one would expect for their age: "looking younger."

[0067] **Fig. 20.** Aging as a differential expression trait. In red: individuals with an expression level higher than one would expect for their age: "apparently older." In blue: individuals with an expression level lower than one would expect for their age: "apparently older."

[0068] **Fig. 21.** Evaluating a delta age for a given gene.

[0069] **Fig. 22.** Model - Principle of aging as a complex expression trait. Left graph: Gene positively associated with age (expression level increasing with age). Center graph: Gene not

associated with age. Right graph: Gene negatively associated with age (expression level decreasing with age).

[0070] **Fig. 23.** Aging as a complex expression trait. Combination across all the genes associated with age for a given individual is achieved by integrating all the genes affected by aging.

[0071] **Fig. 24.** Delta-age in 2 gene expression datasets in a tissue affected by Alzheimer's Disease (prefrontal cortex). AD samples –here used as proxies for accelerated aged samples - display higher Deltas.

[0072] **Fig. 25.** Effect of diet in mice on delta-age (left). Effect of exercise in human muscle on delta-age (right).

[0073] **Fig. 26.** Genetic determinants of aging rate in brain. Transcriptome-wide expression data in brain cortex samples from genotyped, neurodegenerative-diseases free individuals.

[0074] **Fig. 27A-B.** Microglia M1 (Fig. 27A) and M2 (Fig. 27B) gene sets. List of microglial M1 and M2 genes used for Fig. 6B, taken from Supplementary Fig. 10B of Butovsky et al. 2014.

[0075] **Fig. 28.** Genetic variants associated to neurodegenerative diseases association with Differential-aging. Effect of genetic variants associated to Alzheimer's disease, Amyotrophic Lateral Sclerosis, Hippocampal Sclerosis of aging, Parkinson's Disease or Fronto-temporal dementia on Differential-aging value, in a meta-analysis carried in either Discovery+Replication or Disease cohorts as defined in Fig. 3B.

[0076] **Fig. 29.** Strategy overview for identifying TMEM risk variants.

[0077] **Fig. 30.** Local association for TMEM phenocopy in TMEM RR individuals at the IL2RA/L15RA locus.

[0078] **Fig. 31.** Top hits with LD-based proxies.

[0079] **Fig. 32.** Effect of IL2RA genotype on Delta-Age in TMEM106B individuals.

[0080] **Fig. 33.** Effect of TMEM rs1990622 on Delta-Age in the whole cohort, stratified by disease status.

[0081] **Fig. 34.** Effect of IL2RA rs12722515 on Delta-Age in the whole cohort, stratified by disease status and TMEM106B genotype.

[0082] **Fig. 35.** Cross-sectional rate of cognitive decline measured by Mini-Mental Score Examination, stratified by TMEM 1068 and IL2RA genotypes.

- [0083] Fig. 36. Longitudinal rate of temporal atrophy, based on regional MRI measurements at baseline and after 24 months, stratified by TMEM106B and IL2RA genotypes.
- [0084] Fig. 37. Effect of TMEM risk allele, aging and AD on IL2, IL2RA, IL2RB, IL2RG, IL15 and IL15RA.
- [0085] Fig. 38. CNS cell type expression pattern of the identified genes of interest and their ligands.
- [0086] Fig. 39. CNS cell type expression pattern of the identified genes of interest and their ligands.
- [0087] Fig. 40. CNS cell type expression pattern of the identified genes of interest and their ligands.
- [0088] Fig. 41. Genetic modifiers of TMEM106B and the top hits using GWAS.
- [0089] Fig. 42. Shows genes which show a pattern of expression similar to IL15RA in the LPS dataset.
- [0090] Fig. 43. Shows a strategy overview for genetic determinants of pathways of interest.
- [0091] Figs. 44A-O. Shows the top 15 CC GO category decreased by TMEM106B risk allele. An enrichment plot is shown for (A) GO_SYNAPTIC_MEMBRANE (B) GO_POSTSYNAPSE (C) GO_CATION_CHANNEL_COMPLEX (D) GO_POSTSYNAPTIC_MEMBRANE (E) GO_POTASSIUM_CHANNEL_COMPLEX (F) GO_NEURONAL_POSTSYNAPTIC_DENSITY (G) GO_PRESYNAPSE (H) GO_AXON_PART (I) GO_NEURON_PROJECTION_TERMINUS (J) GO_TRANSPORTER_COMPLEX (K) GO_EXOCYTIC_VESICLE_MEMBRANE (L) GO_EXCITATORY_SYNAPSE (M) GO_EXOCYTIC_VESICLE (N) GO_PRESYNAPTIC_MEMBRANE (O) GO_AXON. The top graph shows the enrichment profile, with the y-axis showing the enrichment score (ES), the zero value is represented by a thicker axis line with the axis shown in 0.1 increments. The middle vertical bars correspond to hits. The bottom graph is the ranking metric scores, with the y-axis showing the ranked list metric (PreRanked) with a scale showing values of 5, 0, -5 and -10 and the x-axis showing Rank in Ordered Dataset, with a scale starting at 0 and increasing to 45,000 with the axis shown in 5,000 increments. For example of enrichment plot axis see Fig. 45.

[0092] **Fig. 45.** Enrichment plot: GO_SYNAPTIC_MEMBRANE: Profile of the Running ES Score & Positions of GeneSet Members on the Rank Ordered List. Fig. 45 shows detail of the Synaptic Membrane gene set enrichment.

[0093] **Fig. 46.** Shows genes from the Synaptic gene set downregulated by TMEM106B risk allele underlying the enrichment.

[0094] **Fig. 47.** Shows genome-wide scan for genetic determinants of Synaptic genes levels in human brain.

[0095] **Fig. 48.** Shows the effect of TMEM106B genotype and disease status on aggregated synaptic genes levels.

[0096] **Fig. 49.** Shows the effect of TMEM106B genotype and age on aggregated synaptic genes levels in unaffected.

[0097] **Fig. 50.** Shows the effect of TMEM106B genotype on specific synaptic genes from the gene set.

[0098] **Figs. 51A-L.** Shows the top CC GO categories increased by TMEM106B risk allele. An enrichment plot is shown for (A) GO LYSOSOMAL LUMEN (B) MHC PROTEIN COMPLEX (C) GO VACUOLAR LUMEN (D) GO CELL SUBSTRATE JUNCTION (E) CYTOSOLIC RIBOSOME (F) GO CYTOSOLIC SMALL RIBOSOMAL SUBUNIT (G) GO PROTEIN COMPLEX INVOLVED IN CELL ADHESION (H) GO BASEMENT MEMBRANE (I) GO BASAL PART OF CELL (J) GO ANCHORING JUNCTION (K) GO ACTIN FILAMENT BUNDLE (L) GO MHC CLASS II PROTEIN COMPLEX. The top graph shows the enrichment profile, with the y-axis showing the enrichment score (ES), the zero value is represented by a thicker axis line with the axis shown in 0.1 increments. The middle vertical bars correspond to hits. The bottom graph is the ranking metric scores, with the y-axis showing the ranked list metric (PreRanked) with a scale showing values of 5, 0, -5 and -10 and the x-axis showing Rank in Ordered Dataset, with a scale starting at 0 and increasing to 45,000 with the axis shown in 5,000 increments. For example of enrichment plot axis see Fig. 52.

[0099] **Fig. 52.** Enrichment plot: GO_LYSOSOMAL_LUMEN: Profile of the Running ES Score & Positions of GeneSet Members on the Rank Ordered List; Enrichment plot: GO_MHC_PROTEIN_COMPLEX: Profile of the Running ES Score & Positions of GeneSet Members on the Rank Ordered List. Fig. 52 shows detail of the gene set enrichment.

[00100] **Fig. 53.** Shows genes underlying the enrichment.

DETAILED DESCRIPTION OF THE INVENTION

[00101] The patent and scientific literature referred to herein establishes knowledge that is available to those skilled in the art. The issued patents, applications, and other publications that are cited herein are hereby incorporated by reference to the same extent as if each was specifically and individually indicated to be incorporated by reference.

[00102] The singular forms “a”, “an” and “the” include plural reference unless the context clearly dictates otherwise. The use of the word “a” or “an” when used in conjunction with the term “comprising” in the claims and/or the specification may mean “one,” but it is also consistent with the meaning of “one or more,” “at least one,” and “one or more than one.”

[00103] As used herein the term “about” is used herein to mean approximately, roughly, around, or in the region of. When the term “about” is used in conjunction with a numerical range, it modifies that range by extending the boundaries above and below the numerical values set forth. In general, the term “about” is used herein to modify a numerical value above and below the stated value by a variance of 20 percent up or down (higher or lower).

[00104] The terms “animal,” “subject” and “patient” as used herein includes all members of the animal kingdom including, but not limited to, mammals, animals (*e.g.*, cats, dogs, horses, swine, etc.) and humans. A subject, according to the invention includes, but is not limited to a human.

[00105] The progression of human age-associated traits, such as cognitive decline, is highly variable across the population, with some individuals appearing older or younger than expected at a given chronological age. In humans, age-associated phenotypes such as altered cognition occur at variable rates in healthy individuals (Deary IJ, et al. Genetic contributions to stability and change in intelligence from childhood to old age. *Nature*. 2012 Feb 9;482(7384):212–215). The association between healthy aging and disease remains unclear; for example, brains from health individuals show evidence of CNS amyloid plaques associated with Alzheimer’s disease (Yu L et al. The TMEM106B locus and TDP-43 pathology in older persons without FTL. *Neurology*. 2015 Mar 3;84(9):927–934). Genetic factors have been associated with extreme age phenotypes, such as progeria. Genome-wide association studies (GWAS) have associated complex traits with common genetic variation (Welter D et al., The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res*. 2014 Jan 1;42(D1):D1001–D1006).

[00106] Described herein is an unbiased method for quantifying age-associated individual variability in biological traits, such as gene expression, called differential-aging (Δ -aging). The method can further comprise subsequently performing a genome-wide association study to identify genetic loci associated with differential aging. Differential-aging allows for tissue- or age-range-specific assessment of phenotypes, unlike alternative methods that assume the variable rate will be constant at all age ranges and in all tissues. The method can receive transcriptome data or other biological markers correlated with age as an input and then determine which subset of markers represent variable aging, making the method unbiased and flexible. In some embodiments, the method is used to analyze transcriptome-wide cerebral cortex gene expression. The method assesses an individual's biological age based on the biomarkers and determines if the computed age differs from true chronological age. This results in the identification of genetic variants related to aging.

[00107] Also described herein is the identification of the TMEM106B gene locus as a determinant of age-associated changes related to the brain. TMEM106B was identified using the differential-aging method followed by a genome-wide association study in search of genetic modifiers of Δ -aging. The method used cerebral cortex gene expression data to show that the rate of human cortical aging depends on the TMEM106B gene, a gene previously associated with frontotemporal dementia. Therapeutics can be developed which target TMEM106B for the treatment of age-related cognitive decline. The TMEM106B gene locus was identified as a determinant of Δ -aging in cerebral cortex with genome-wide significance ($p < 10^{-20}$), in a meta-analysis of several cohorts totaling 1904 autopsied human brain samples. TMEM106B risk variants promote age-associated changes, such as inflammation, neuronal loss, and cognitive deficits, even in the absence of known brain disease. Surprisingly, the effect of the TMEM106B risk allele on Δ -aging is highly selective for the frontal cerebral cortex of older individuals (>65 yo). These data suggest a mechanistic link between accelerated brain aging and neurodegeneration.

[00108] *Methods of determining biological age and differential aging*

[00109] In certain aspects, the invention provides a computer-implemented method of determining a biological age of a sample from a subject comprising: a) providing a gene expression level of a plurality of genes in a sample for a plurality of individuals; b) performing a linear regression of the expression level of each gene of the plurality of genes as a function of chronological age of each of the plurality of individuals to determine a linear

regression co-efficient for each gene of the plurality of genes; c) selecting, from the plurality of genes, the genes whose expression level is significantly correlated with chronological age; d) determining, for each gene whose expression level is significantly correlated with chronological age selected in step (c), a residual value, wherein the residual value is determined by comparing the gene expression level of said gene in the sample from the subject to the linear regression performed in step (b) for said gene; e) determining, for each gene whose expression level is significantly correlated with chronological age, a ratio of the residual value in step (d) to the linear regression coefficient in step (b); f) repeating steps (d) and (e) for each gene whose expression level is significantly correlated with chronological age selected in step (c); and g) integrating the ratios for each gene whose expression level is significantly correlated with chronological age, wherein the integrated ratio corresponds to the biological age of the sample from the subject.

[00110] In some embodiments, the method further comprises comparing the biological age to a chronological age of the subject to determine a differential aging value for the subject. In some embodiments, the plurality of individuals includes the subject. In some embodiments, steps (d) to (g) are performed for each individual in the plurality of individuals.

[00111] In some embodiments, the method further comprises performing a genome-wide association study (GWAS). GWAS is an observational study of a genome-wide set of genetic variants in different individuals to identify variants associated with a trait. GWAS approaches often focus on associations between single-nucleotide polymorphisms (SNPs) and a trait of interest. A SNP is a variation in a single nucleotide that occurs at a specific position in the genome. SNPs often underlie differences in susceptibility to diseases. For example, SNPs cause a wide range of human diseases. SNPs can also affect the severity of illness and risk levels for certain diseases. For example, a SNP in the apolipoprotein E (APOE) gene is associated with a higher risk for Alzheimer's disease. GWAS studies compare the DNA or genomes of individuals having varying phenotypes for a particular trait or disease. These individuals may be people with a disease, and similar people without the disease, or they may be people with different phenotypes for a particular trait. For each individual a sample of their DNA is provided, from which millions of genetic variants are read using SNP arrays. If one type of the variant (one allele) is more frequent in people with the disease or people with a particular phenotype or trait, the variant is said to be associated with the disease or with the particular phenotype or trait. The associated SNPs are then considered to mark a region of the

human genome that may influence the risk of the disease whether an individual has a particular phenotype or trait.

[00112] In some embodiments, the GWAS identifies single-nucleotide polymorphism (SNP) modifiers of the differential aging value in the plurality of individuals. For example, as described herein, GWAS studies can compare the DNA or genomes of individuals having varying Δ -aging values. The variants more frequently associated with high or low Δ -aging values can be identified. These associated SNPs identify genes and alleles that influence Δ -aging values and thus biological aging in a subject.

[00113] In some embodiments, the GWAS identifies genetic modifiers of the differential aging value in the plurality of individuals. Genetic modifiers can include genes whose expression levels are correlated with the differential aging value in the plurality of individuals.

[00114] In some embodiments, the gene expression level is the gene expression level in the brain. In some embodiments, the gene expression level is the gene expression level in the frontal cortex. In some embodiments, the gene expression level is the gene expression level in the cerebellum.

[00115] In certain aspects, the invention provides a computer-implemented method of determining a biological age of a sample from a subject comprising: a) providing a gene expression level of a plurality of genes in a sample from a subject; b) determining, for each gene whose expression level is significantly correlated with chronological age, a residual value, wherein the residual value is determined by comparing the gene expression level of said gene in the sample from the subject to a linear regression of the expression level of the gene as a function of chronological age for a plurality of individuals; c) determining, for each gene whose expression level is significantly correlated with chronological age, a ratio of the residual value determined in step b) to the linear regression coefficient for the gene; d) repeating steps b) and c) for each gene whose expression level is significantly correlated with chronological age; and e) integrating the ratios for each gene whose expression level is significantly correlated with chronological age, wherein the integrated ratio corresponds to the biological age of the sample from the subject. In some embodiments, the genes whose expression levels are significantly correlated with chronological age can be determined according to the methods disclosed herein.

[00116] In some embodiments, the genes whose expression was significantly correlated with chronological age are PNOC; TAC3; FKBP5; CRH; ITGB4; VIP; PVRL3; GMPR; C1QTNF4; ITPKB; RARRES3; GRIK1; NPPA; NMU; HBQ1; ACOT7; PENK; OLFM1; BZW2; ARPC2; ADRA2A; CA10; SSTR1; SLC24A3; LGI2; ACTR3B; MTMR11; CYP46A1; ARHGEF6; KCNF1; PTPRR; NPC2; RNASE4; ARHGAP20; PDCD2; CALB1; MAOB; CX3CL1; OTOS; NRXN1; LRRC1; SERPINA3; GNG3; NEK2; SCN3B; NFKBIE; SLC10A4; TTL; FBXO9; FGF13; CAMK4; STMN1; SLC32A1; ING2; RFC3; PSMD7; STYK1; DNAJB5; BAD; RIMS1; HSF4; SRD5A1; RPL39L; GRP; GPR26; SLC16A9; ACP1; FAM5C; OGG1; AQP1; DCUN1D5; RUFY1; NRXN3; AZGP1; SRRM2; PIN1; FLRT1; BCL11A; UQCRRF1; CORT; VWF; TCEB1; DLX1; HAPLN1; TPP1; PTHLH; TRAF1; MRPL28; UBE2E3; EEF2K; HDAC9; TAC1; TASP1; CRYM; GNB5; ERBB2IP; MAL2; WDR19; COPS4; DOK6; CXCL14; SERINC3; DHRS3; FGF12; GAD1; PRDX1; TRAM1; CAPS; GALM; TNFAIP8L1; ATP6V0D1; CHCHD6; SLC14A1; NCALD; PHF14; NDN; PRKCZ; PHKA2; RPH3A; GRASP; CYCS; SVOP; STXBP5; RPS27L; FUK; SEC61A2; DGKI; BMP4; FBXW9; MMD; MAD2L1BP; ACADVL; C1S; TRIP6; ATCAY; ABI2; CD44; AP1G2; SLC22A5; LAMP2; HOMER1; TSSC1; GPR61; PRKG2; SMPD3; TPM3; PSTPIP1; YWHAZ; AEBP1; PRKX; HEBP2; RELN; MFSD4; KCNK3; MS4A6A; NXPH1; DOK5; TUBB2A; RHOBTB3; ZBTB16; SIRT5; TMEM38A; RAPIGDS1; LARGE; ATP2A2; PLSCR4; CACNA1G; CKAP4; DNAJB2; TLN1; TIMM17A; MAGI1; RQCD1; KIF3B; CSRP1; FBXW7; FCGBP; GRAMD3; FAM49A; CXADR; CDC42; CRYGS; HPS5; STAT2; PITPNA; SCGB1D2; GLRA2; RAB3C; FYCO1; RERE; PLEKHB2; MED8; CECR6; SSBP4; HSPBAP1; PANK3; SETDB1; EPM2AIP1; CCNA1; CALB2; TRPV1; AHCYL1; ATP5J; CD200; VLDLR; EEF1B2; TRPC3; EFHD1; BDNF; CBFB; TAZ; MKLN1; GSN; DDR1; FAM19A2; DLL3; RIT2; MGAT4B; MT1F; IL17RB; TUBB3; SYTL4; SPATS2; AP1S1; GYPE; TYK2; STK4; GAD2; VKORC1L1; NOL4; PSRC1; FGF14; AFF1; LTBP3; LRRTM4; ADH1A; BCL2; EBP; IGF1; PLOD1; SLC39A3; ANTXR2; CDC42EP3; TRIM9; DCX; PHKG1; GPM6A; GRIN3A; PPFIA1; MRPS7; COX5A; IDS; ARL4D; PAK7; TJP2; ATP6V1G2; SLC25A28; MGLL; SPR; TRIM39; SLC39A1; MRPS12; NDUFA9; HN1; CNTN4; ATXN10; MRPL37; MAN1A1; CCNC; GSG1; LRRC7; ID3; MYO5A; GABRB3; MED19; CBS; ZMAT2; HRASLS; ELMOD1; SCG2; CDK5RAP3; MN1; RGS12; NDUFS4; IL13RA1; RCN2; MTIX; TXNIP; FAM53B; RASGRF2; MYOT; COX8A; IFITM3; CSNK2A1; CDKN3; LHX6; CAPG; DCTN3; BEX1; RASGRP1; FKBP1A; MRPS23; NPHP3; GSTA4; FBLN2; PLXNB1; PGAM1; NPM3; MRPS33; CNTNAP5; LGMN; SHMT1; NUP214; HTRA2; PCBP2; IFITM2; LRPAP1;

CDK5R2; BAZ2B; HSPA8; CLK1; DGKB; PAFAH1B1; LBH; PDE8B; ELF2; TOR1AIP1; CACNB1; DLX2; KCNG1; STAT3; DUSP14; UBE2V2; LSS; RGS6; FGF9; BIN1; ZC3HAV1; CDKN2C; MRPL14; PRSS16; ERCC1; CCND2; NPY5R; EML3; BCL2L13; PTPRO; SORBS1; ST8SIA5; PKIG; SLC6A8; C1RL; CYP27A1; RASL12; SUB1; GRIK2; KAZALD1; ZAK; RPS6KA2; NLGN4X; MT1G; BBOX1; GUK1; CFH; GNA13; CDH9; SAFB2; SLC25A4; ANG; NDUFB6; CBLN1; PILRB; KHDRBS2; SNX3; NBEA; MFSD3; RPS6KA3; CDKN2D; CYP2J2; ALDH1L1; PIP5K1B; RFX2; NNAT; VIM; GPHN; MACF1; ADAM23; UCHL3; PDGFRA; HSPBP1; TM2D3; LMO3; MBD6; DLG3; DKC1; RTN4R; SP8; R3HDM1; KCNAB1; PTGES2; TGFBR3; ADH1B; CNOT7; VSIG4; GAB1; IQGAP1; ECHDC2; SPATA18; PARP9; KRTCAP2; SPEN; ANXA5; DYSF; CYP26A1; LMBR1; MAP2; NUMA1; CDK8; ACO1; SLC16A2; KCNN3; CDK5; RHOQ; BSDC1; RRAGB; PITPNB; CRYAB; CAMSAP1; DUSP6; PHGDH; PIAS2; SOX4; SH3BP4; LRRFIP2; CLEC4G; BAZ1B; SMOX; ANXA4; SORBS3; PTS; DLAT; UBE2D2; VPS29; ATP2B1; ACTR10; PRKCD; CPS1; ZFYVE26; RAB3A; SLC9A9; LPIN1; TIMM13; SLC24A6; CLCA4; TIMP2; TMEM66; AQP4; LMO4; PMM1; PRPS1; IARS; KLF15; EFNA1; FAM3C; CPT1B; BCAP29; PRELP; SIPA1L2; PPP1R14C; GPRC5B; SAV1; TJP1; ARL4C; CAPN2; NHP2L1; CA5B; WSB2; ADIPOR2; CDH8; BBS1; APEH; MDM1; SIAH2; TBC1D7; CRKL; ZNF277; PTGDS; MICA; GABRA5; FNBP1; SET; CRMP1; MAZ; GOSR2; STRAP; HIF3A; B4GALT4; GSTK1; CHD4; FTCD; FANCG; ATP5F1; CBLN4; MVP; SORT1; NEUROD6; STOM; PLTP; RAPGEF6; LCAT; PLD2; DDIT4; CRYL1; RPS6KA1; HSPA2; PSAT1; THOC1; PCMT1; IFIT2; SMPD2; KCNJ2; SLCO2A1; ENSA; WDR17; MED12; AP3S1; PDK4; CACNB2; NME1; HS6ST2; ADD1; ZCCHC17; MYB; SNAPC5; SLC27A3; PKD2; UBE2V1; SLC30A3; JUP; ARHGEF7; SLC38A2; RNF130; CEP68; TPD52L1; HOMER2; IL13RA2; VAMP5; DNMI1; NPTX2; RTN2; CCNDBP1; ACTL6A; H2AFJ; BRE; CAMK2N1; MRPL3; SARS; AHNAK; GPM6B; LDOC1L; NECAP2; POGZ; BAI3; XRN1; HCLS1; GUCA1A; PLEKHA4; CGREF1; TENC1; ACACB; MAN2C1; MCTS1; SGIP1; ZXDC; TBL1X; DYNLL2; KCNC2; GPT2; LPL; PSMA1; CRHBP; HIVEP2; DTNA; HES4; SYT17; GPBP1L1; GLTSCR2; STAT5A; ZNF329; SCRIB; IRF3; DLGAP4; SWAP70; PJA1; KPNA2; DCLRE1C; CASP6; PWP1; TUSC3; NELL1; C7; SNTB1; S100A10; RAB3IL1; MARCKS; WFS1; FAHD1; PPP1R13L; BAZ1A; LANCL2; LEPR; NME7; SUCLA2; RYR1; KCNK1; MTIM; MEGF10; TXN; FLCN; TTC18; TNFRSF10B; UCN; SLC15A4; RDH12; RTN1; SLC25A1; PILRA; MGST2; CP; NDUFB2; GABRG3; HMGCS1; PRDX6; ZNF395; RTN4RL1; ATP5A1; NPL; ACSS2; PARP11; HSD11B2; FNDC4; ARRDC4; SLC25A5;

KCNV1; PARP6; ARHGEF10; MX1; LRIG1; CPSF1; SNURF; MRPL4; PLD3; LZTS2;
 NKTR; MYO1B; PPP1R7; DIO2; RAF1; TMPRSS3; CASC3; SEC61G; HINT1; PBXIP1;
 PCDH10; BGN; CHRNA4; PLEKHB1; KCNH7; DYRK2; TBC1D2B; STX1A; SLC25A3;
 POP7; FZD7; CTDSP2; SLC37A3; NEBL; CIAPIN1; PMP22; ANKRD39; MATN2;
 LAMB1; DOCK1; EGR1; FGFR1; LSM3; GPC4; MAP1LC3A; SCAND1; UBE2T;
 MGST1; PRODH; COL4A5; TFAM; ISLR; BCL6; SMYD3; PHACTR4; DNALI1; FNDC5;
 CIZ1; RFX1; GYPC; IFITM1; HAPLN2; NAP1L1; SNX25; FHL2; HSD3B7; CACNA1H;
 PIGN; BCAR3; COMMD9; TPI1; SEPHS2; DSTN; CDCA5; PLOD3; ATP2B2; DLX5;
 MRPL33; ANKFY1; LIMK2; NSF; HMP19; CAMKK2; OTUB2; RYR3; NPC1; APC;
 MAGEH1; METTL6; CHST3; SNRPD3; STEAP3; ZNF3; AADAT; ITGB1; PRSS23;
 SLC39A11; MAPK1; SNCA; NDUFA4; SHANK2; GALC; ATP6AP2; GTDC1; TMEM43;
 VAMP8; ENTPD3; RAB40C; LRRC20; HMGCR; ROM1; COMMD7; AKAP12; VPS25;
 PLCD1; LGI4; OXCT1; RBM12; MAP3K6; PGAM2; ILK; FEN1; NLE1; NDUFB3;
 ACVR2A; RARRES2; ST6GALNAC3; NGFRAP1; CSMD1; PDGFRL; ST3GAL2;
 FBXO3; GUSB; TACC2; ZNF436; FSTL3; CLASP1; SEH1L; KCMF1; SELL; CRTAP;
 TRIM56; CAMK1; PCDHB3; ZFAND2A; KCNS3; FCHSD1; RAB2B; MTF1; FGR;
 TNFRSF18; POLR3K; CHCHD2; MRPS17; PORCN; PPP2R5C; PC; CAST; FABP3;
 SRPX; CDK5R1; RAPGEF3; PLCB1; LYPLA1; PDE4D; RTKN; RABL5; NCAM2;
 RGS16; HEPH; BUB3; NFKB2; ST5; FEZ1; LMNA; MDH2; GNRH1; LGALS3BP; SOX9;
 WBP11; ASCC2; IGF2R; SPATA7; KLHDC3; TTC21A; HERC5; NXF1; HSPB8; AQP11;
 WTAP; SMAD4; CGN; WDR49; COL9A1; KCNMB2; KLKB1; POMC; SIN3A; PRPS2;
 PAK1; MANBAL; CLSTN3; PPP3CB; PHLDA3; SLBP; CNTN6; ZC3H8; ABHD4; JAZF1;
 BCAT2; GRB2; TNFAIP1; G3BP2; ERBB3; PPP1CA; KATNAL1; UBE2Q1; MYL3;
 PGK1; MYO9B; NFKB1; ATP5G3; MRPL45; RABEPK; APBA3; PSMA6; NUDT15;
 TJAP1; 42066; NRN1; PYGB; FCGR2A; CDH12; HSD17B4; ANKRD49; CDH22; RPRM;
 LMTK2; NDUFS3; APOL3; CRYZL1; TRAPPC4; MKKS; SPG20; NFATC3; POMT2;
 DVL2; PSMB8; DAB1; SYNGR3; KIF13B; PPM1E; ARIH1; STRBP; TMLHE; SMPD1;
 MAOA; RPL41; NTNG1; EHD1; SLC03A1; TIMP1; TPD52L2; CDS2; VRK3; COL7A1;
 ATRN; CACNG3; SRPK2; CDC40; GRM7; PDCD10; SEMA3B; UCKL1; VGF; TMEFF1;
 EXOC6; HEY1; CHDH; ACIN1; MAP2K4; BCAS4; MAPK9; MAPK3; MRPS35;
 LONRF2; SLC39A4; APOD; BBS2; SNAP23; ADD2; ZHX3; PAK2; SMAP1; PRKACB;
 PAIP2; RGS7; EMP3; NSDHL; PLEKHA7; PPIH; FYN; SNTB2; TSFM; RPGR; VTI1B;
 INPP5F; GRAMD1C; MUC1; TPST2; MRPL46; CLTA; ZRANB1; RHEB; COX6C;
 TNFRSF11B; EPHA5; PDHX; KCNIP4; PRRT1; CHD9; OSBPL9; SMAD3; OSBPL11;

POLE; ANKRD35; CDK7; PPP2R2B; RFX4; KCTD8; GOLPH3L; PPCS; DYNC2LI1; CALCOCO1; LHFPL2; ATG7; KDELC2; UBE2G1; SLC12A9; HLA-DMA; RBM18; CDH19; ADCY1; PIM2; RASEF; ABCA1; MRPL55; TMEM39A; SMAD5; CDKN1C; ADRB1; PDCD6; CAP2; CD163; UCK2; GNAL; WDFY2; SLC15A3; DPF2; ARL5A; VAMP2; NECAP1; ADK; LIFR; NETO1; SLC35B1; WDR20; TUBG1; RAB15; HDAC1; PSMA5; CWF19L2; HECTD2; PDRG1; PCDH8; LRRTM1; METTL1; SPOCK1; ATP5G1; GHITM; MOGAT1; QTRT1; CCKBR; ARAF; DPP7; STARD3; OR7A5; RALB; RAD51C; EGR2; PPP2CA; CHI3L2; ARFIP2; RNF135; SSH3; UROS; TIMM8A; EGR3; NEO1; CTNNA2; HMGA1; NDRG3; ITPK1; PITPNM2; HLA-DMB; QDPR; B3GALT6; ITGA2B; SNAP25; MTM1; POLE2; CAV2; NFS1; USP47; RIOK1; RGS4; ADD3; MAT2A; NKX2-2; LPP; LCMT1; ATP1F1; SUMO2; TFB2M; IL20RA; TFAP4; MLH3; RUNX1T1; EPHA4; SLC26A11; SFXN1; SLC11A2; MAEL; CHCHD3; YAP1; EIF2AK2; GPR21; SLC29A1; UBE4B; CHRNA7; ARHGAP9; NEDD8; UST; NAV3; PCDHA3; ZNF226; APH1B; MRPL1; MAP3K13; MID1IP1; GABBR1; MICB; CCND3; MAPKAPK5; DDX25; RNF175; PUS1; RASGRF1; LAMC1; NPAS1; ATP6V0C; HLA-DPA1; INPP5A; STMN2; ERBB2; NDUFB5; ACAT2; FJX1; DMXL2; CD86; CFLAR; SAE1; SLC25A25; E2F5; PRDX5; CITED1; SYP; SLC30A9; DNASE2; DDX10; ELAVL2; DPYSL3; LEMD2; SPTBN1; SMUG1; LRRC8A; COPZ2; LITAF; DHDDS; UBE2W; ARMC8; FMNL2; ZNF692; TCEAL7; RAN; YWHAB; CNKSR3; CDA; CHL1; MCM2; ACOX2; HDC; PACSIN2; ANXA2; COMMD8; HSPB3; PYGL; MT2A; SLIT1; WASF2; SUPT6H; SQLE; NCOA4; CHGB; RND1; TRIM22; IDH3B; TRUB1; NLK; EDN3; TM9SF1; GNA12; MAGED1; GSS; CREB5; SNRPA; EXOSC5; RBM6; NRIP2; HRAS; PSAP; PCP4; ASNS; GJA1; CCNH; PTPRN2; WDR6; HS6ST3; SLC30A7; DBI; PPP2R1B; OGFR; SNX17; PHF1; EEF1D; ANP32B; SDC4; ZNF217; UACA; AGGF1; PON2; RPAP1; NKIRAS1; RHOG; NDUFAB1; BTN3A1; LGR6; PKD1L2; SEC24C; GABRA1; CCDC28B; GDAP1; FUT9; CGNL1; GDA; AGT; ACTN2; AKT1; MTFR1; DYRK1A; SREBF1; ME3; RENBP; STX12; ELF1; CREB3L2; CEP135; PTTG1; CCPG1; UNG; MGEA5; LMNB2; CBLB; PTPRU; TTC19; GSTP1; IFT140; RAP2B; TOB2; PRR7; SLC7A2; EEF1E1; CCT6A; KIF7; PFKFB2; SUSD1; PLCL2; EIF4E; RHOA; TNPO1; UPF3B; PANK1; UQCRC2; KCNK4; SPA17; USP46; CBX6; MRPS30; BOK; CBX4; MRPL36; UHRF1; RFK; OMA1; GAP43; AKAP13; CD22; ACVRL1; TRADD; CDO1; OCA2; SIX5; MCM3; PIAS3; MPPED2; EVI2B; EMX2; PCNXL2; LZTR1; SLC25A15; ST8SIA3; MT1E; EWSR1; ABCC5; CLK3; ACBD6; E2F3; TPR; FAM5B; SLC44A3; IL15; DNAJC18; PARD6A; ANGPT1; SCARB2; RHBDF1; TRAF5; MT1H; SQRDL; ARF4; RPUSD3; MAPT;

GDPD2; ANXA1; RXRA; SCO1; DUSP2; ADRA1D; TNRC6A; SERPINF1; F8; SASH1; TST; CHST6; ASXL1; SERPINB2; NKX6-2; TMCO4; CACNB3; NRG1; PDGFRB; CCM2; ZNF384; PLXNB3; PMP2; COX10; DUSP3; TCEAL1; ATHL1; PCDH7; MFAP1; AP3B1; GPR17; DENND2D; GMNN; ZNF626; CTDSPL2; PTPRN; BNIP2; METAP1; AK2; SFRP1; SEZ6L; ALDH2; CRYGD; CA11; RARB; EBNA1BP2; DDR2; ZMYM1; TSPAN17; PPP1R3D; MAPK10; PSD4; SUPT4H1; PCF11; CDKN2A; DHX36; S100A11; RUVBL2; SEZ6L2; ERP29; CLDN15; COL21A1; EYA2; TAF9; TP53BP1; CHST13; STAT5B; LRP10; SNX5; TFAP2E; RBL2; PPP1R10; COL24A1; RTN4; TNFRSF25; CAPN3; KIAA0355; FXYD5; CRB2; ST8SIA2; NSUN6; TMPRSS5; RSNB1; DPM1; PANK2; FRMD4A; CNKSR2; TMED3; FXYD1; SLN; ABLIM1; NFKBIA; ARHGEF1; ITGA6; ZIC2; TRHDE; ZNF175; ALPK1; DPM2; SELENBP1; TCP1; ING3; CLIC4; TLR2; NGEF; SYNCRIP; FCN3; SMS; SSPN; CHMP5; SLC4A11; STAG1; SOX13; SLC25A20; ST6GALNAC5; HSPA4; NSUN5; UCHL5; TUSC2; CTNNAL1; IFI35; CAPN9; FGF2; SSFA2; CHRFA7A; RFC4; RAB31; CCT2; SPTLC1; SYT1; PPP1R9B; RPL23A; RALGDS; EHMT1; CD300LB; PABPN1; STRC; NDFIP1; RNASET2; ISG20; SOCS2; ARPC5L; SCRNB3; TNFSF4; ANKMY2; TUBGCP6; ARHGEF17; CTNNB1; MTX2; KIAA0195; RAPGEF1; CBR3; ARRB1; PHF21A; CYP4F11; CTPS2; B4GALT3; BMPR2; ACTR5; ARMCX3; CTTN; NCSTN; GIMAP1; AP2S1; GREM2; PPM1B; MRPL12; NFATC1; CMAS; PHB; DPY19L4; PRRG1; TMCC2; PPP1R1A; PKN1; IFT88; ZNF324; VPS13C; TMEM2; GRK5; MAPRE1; LRRN1; DRP2; DST; CPNE6; PPIL1; SLC44A2; FCER1G; PAICS; ZNF224; NNT; ABCA8; ZNF667; RPL19; SSSCA1; KCNQ5; FMO4; EDN1; GDPD3; YPEL3; ZNF573; UQCRCQ; ZNHIT3; SENP5; CNOT10; ACTG1; PAK6; LMAN2L; PRSS3; DOC2A; TROVE2; HTR2C; RFWD2; RBP4; KRAS; RASA3; RNF11; S100A9; CLU; IFI16; N4BP1; SCAMP1; SMO; SLCO4A1; APBB1; GRIA1; TXNL4A; TNIP2; IL17RC; PCSK1; PSMB6; DYNC1LI1; FSD1L; CALD1; SLC25A18; CLMN; LRRC48; NOTCH1; MMRN2; DUSP12; SCAMP2; DNM2; KIT; CYC1; BAI2; NADSYN1; LYSMD2; FKBP3; TBRG1; EPS8; DLGAP1; NEK3; CYP26B1; PDE1A; LNX1; WNT7B; NTRK3; SERPIN1; CCNI; KIRREL2; OCIAD2; BAG3; WIPI1; ELP4; RPS6; SLC35C1; AATF; USP37; VBPI; UCHL1; DOCK5; CAV1; YTHDC2; MAN2A2; RBM17; SDK1; CLCN7; NDUFS6; ACADL; LRP8; AACS; CD81; ALDH6A1; PRTN3; PTPN13; XPO7; USP54; LY6G5C; VPS41; YPEL1; ADCYAP1; CRK; HIST1H1C; SERPINF2; MIER1; PDLIM3; RPS6KB1; FAM49B; KRT5; ADAM15; TTC4; CD14; NBL1; ZNF606; FIBP; RAB11FIP2; FAM19A5; ARL15; WEE1; NPEPPS; ARX; TBC1D19; CUEDC2; IRF2BP2; GNL3; FANCC; PTPN3; SENP3; ST3GAL5;

GABARAPL1; PLA1A; THBS4; LDHB; CDH1; TMEM14A; VDAC1; APLN; KCNH2;
PCDHAC2; CELSR3; NRIP3; CTDSP1; NIPSNAP1; TLE4; CDC42SE2; HCN1; GNPTAB;
XKR6; ETV4; HELZ; SLC26A6; DLL1; AASDH; EHP1; SYCE1; RGS14; HSBP1;
EPOR; CHUK; DFFB; GAS1; GLRB; ZNF189; ALDH3A2; SLC25A27; RBBP8; KCNK2;
CLN3; TEX14; KCNQ2; CEP164; GSTM3; SLITRK1; DYNC1H1; STARD13; SLC4A2;
RNF128; SLC44A1; TNFRSF1A; WDFY1; NUDT18; KIF3A; DHRS1; MAGEF1;
NDUFC2; NDRG4; PRAM1; DCK; BCLAF1; ZDHHC3; DDX54; NOL3; RPL34; PSMA4;
ACTA1; TCF12; TAGLN2; CACNB4; NDUFA11; PSMB7; RASSF4; ST3GAL3; TRIB3;
HIST3H2A; ECH1; STK36; IL4R; LSM4; DPH2; RTN4RL2; SDHB; NUBP1; PHF8;
ETV5; ZNF451; RNF14; ARHGAP17; EXOSC6; ALDH9A1; TRMT1; ARRDC2;
SEC61A1; RHOC; DISP1; EPB41L4B; BFAR; PEX11A; ACSL5; FSTL5; USP7; KLHL8;
MXRA7; SPON2; PSMC1; SEC23A; HDAC6; MRPL15; CPNE3; SGCZ; HPCA;
PPP1R1B; OSBPL6; SCARA3; MRPL54; ZFYVE21; FCGRT; SLC31A2; TOMM22;
PRDM10; YIPF6; PSMC6; LRRN3; RASAL1; SPON1; GLIS3; TTC23; MYO18A; HBG1;
ADAMTS8; MAPKAP1; GSR; ANGPT2; PROS1; CHKA; AHCY; NEDD4; DERA;
COPS7A; EIF5B; ZNF146; TMEM14B; GATAD1; WDR75; IPO9; UTP11L; CNM3;
PSMD8; YES1; SORBS2; CACNA2D1; TNFAIP2; PCMTD1; FAM98C; IVNS1ABP;
LTA4H; RAB3B; RPL7L1; DARS; CBR4; OFD1; GPR176; PLK2; INPPL1; C5; HMGB3;
LIPE; ASCL1; RPS26; ACY3; MEF2A; CHCHD1; HIBADH; WHSC1; ZBTB20; DGCR2;
MT1A; PPAP2C; AHS2; SYNGR4; TNK2; MAML1; ANKRD10; PCDH11Y; ZNF710;
B2M; MRPL11; MYO7A; ATP6V1A; EIF5A2; RNF103; PREP; IGSF3; SPAG1; SFT2D3;
TEKT3; CNIH2; MRPL35; ST6GALNAC2; CREG2; AGTRAP; ST7; GMPR2; PLEKHF1;
CLIC1; CD109; MATR3; MRPL20; CSF3R; NDUFB1; STK25; TLR5; MAP2K1;
ZBTB80S; PROM2; CNTN1; TAGLN3; ANKRD46; LEPRE1; SSX2IP; ALS2CL;
CHAF1B; KLHL24; GLUD1; GBP2; CDKN2B; FAM92A1; BBX; PEX11B; TCF7L2;
CAMKV; ABCA2; NIPBL; PLOD2; NFE2L2; FLNB; RNF5; MEF2D; NRG1; SELT;
TNFRSF6B; KCNT2; DLD; SLC2A14; MAGED2; ZNF544; BTG1; ELL; HTR2A;
RARRES1; BRD4; NUDT22; EPHA3; STARD3NL; GAS6; ARPC4; MXI1; PDK3; ETV1;
LICAM; CSPG5; CIDEA; IFNGR1; KITLG; PGM2L1; TRAF3IP2; TMEM31; ECHDC1;
MRPL30; PIGM; TGM2; LLGL2; RAB10; PCDHA12; LCP1; TSC22D4; SMOC1; MAF1;
ZBTB33; MAP3K5; SPRY4; FADS3; TRIM47; CLN5; HLA-DPB1; AUH; PCBD1; TEPI;
ACSL1; BAIAP3; NICN1; CIC; PPARGC1A; ENY2; NOVA1; HAVCR2; GPR3; SNRK;
CDK3; OMD; TIMM23; USP10; SLC25A37; OTOF; FANCE; PELI2; KCNG3; PIGS;
PSMC4; NOG; TBC1D10A; APOBEC3B; SYT5; TCEA2; HPRT1; AAAS; CTRL; SRRM1;

MRPL16; ZZZ3; HIVEP1; HLA-B; HK1; MANEAL; SEPN1; OGFRL1; SNRPD1; AASS;
DYNLRB2; ILF2; LYPLA2; FBXO2; DHX32; TUBE1; WDR61; PPP1R11; ANAPC7;
PDIA2; NPM2; BYSL; ITGB5; SHC4; PURA; WASF1; USP12; MADD; HEYL; PTPRF;
ZNF529; SDCCAG3; CD74; LRRC8B; ZFP36L1; TSPAN3; PPP2R2C; ADARB1;
PSMD13; AZIN1; MRPL47; TNS1; MAMDC2; PCBD2; SLITRK5; RANBP3; PMPCB;
GTF2F1; LILRB3; ATXN1; TACC1; PINK1; CAND1; ST18; SEMA6B; LIG4; ANKRD9;
HSD17B1; LYZL4; STAB1; PLAGL2; RUVBL1; ACSL4; TRPC6; PKN2; MLLT11;
TGDS; COQ2; ECM2; PCTP; RNF4; RACGAP1; ATP5L; ZNF471; COL23A1; PSMD2;
DAAM2; NCOA1; HSPD1; MBD1; RCC2; FBN3; ZFH4; C3AR1; H2AFY; OSBPL10;
PRKCSH; ABAT; SRGAP1; CD40; BHLHB9; LRRC32; CNTNAP2; KCNJ12; SLC1A6;
PSD2; TM6SF1; MYH9; SERPINA1; CDC42BPA; BCAN; ANXA3; STAMBPL1; ASH2L;
PACRG; COX7A2; AKR1B1; HTR5A; MAP1B; KHDRBS1; WDR77; LEPROT; SGSH;
PPIL4; PCDHA11; MOV10; SPTBN5; DACT1; PTPN6; CNN3; NINJ1; CALM1; CEBPB;
SLITRK4; CERK; NEGR1; IMP4; MAP7; SNAPC4; MAPK14; DPYSL2; NR4A1;
CCRN4L; TSPAN32; TM9SF2; NUDT6; SYTL1; UPP1; STAT6; CHD1; MCM5; 42064;
GAK; SNRPB2; TBC1D14; CD58; CDH10; AMPH; PLCE1; MESP1; PI16; POU5F1;
STK11; MORC4; NONO; UBL7; MAEA; XAB2; ANGPTL4; GALNT9; MRPS34; NIN;
POU3F3; CNOT6; EIF4EBP2; TTC9C; TSPAN13; PITPNC1; TOR1A; BNIP3; NFIA;
ANLN; B3GAT1; ARF5; TMEM63A; S100A6; ALDH4A1; ZBTB40; CLTB; UQCRC1;
TRIM17; CCDC34; RNASE2; HIF1A; SYNGR1; SIGLEC8; SLC20A2; DNAJA1; XPC;
RDX; DIRAS2; PLCB4; PDCD4; PGAP1; GALNT14; TYRP1; ZNF330; GP9; MYO10;
CNTN2; SSR1; NBPFF3; TGM1; SORCS1; TUBB6; POLDIP3; LYAR; SFPQ; TIMM8B;
PTPRA; CNR1; HS6ST1; RANBP17; LRP2; PCGF5; PPP4R1; ZNF34; BCAS1; YWHAG;
TAF4; KCNJ11; TIMELESS; UBE2N; FNBP4; MCM8; COPS3; SOX5; WDR47; ZNF503;
XPO4; UNC5CL; WSB1; ZNF491; C1QB; HOMER3; PXX; SYNGR2; TRIM38; DQX1;
FGF17; RNF20; ZC3H11A; CAD; MECR; KHK; PYCRL; MRPL9; SPSB2; MAL;
ARID4A; LRRTM3; BTN2A1; PLCG1; GNAZ; VWA1; KCNA1; NEU3; ZMYND12;
COPS5; HNMT; COX7B; STXBP6; SLC2A13; ICMT; ZCWPW1; AKR1C3; ITPKA;
ZBTB39; RPL36; SSH2; IMPA1; STAU2; PPM1A; FBXL7; TSNAX; DPP9; AKAP8L;
PEA15; TMEM47; TAF1C; PFN2; SUPV3L1; DPYSL4; SMARCA1; ZNF558; PTK2;
LY6E; NUDT4; NEXN; PHF15; BTBD3; MTMR3; SLC35F3; HSD11B1; SMARCC1;
KIF1B; POLR2D; CHML; ST6GAL2; UXT; GLT8D1; PRKCH; CDC14A; ITPR2;
OSBPL2; PTPN4; TMEM59; GORASP1; ITPR3; CYBB; IGFBP7; STK32C; GNB2L1;
TRAM2; MARK1; LPHN3; SPI; PLEKHA1; KCTD9; NEDD1; VTN; MAFG; RTN3;

RING1; TLR4; ACP5; ABCG1; GIMAP4; UNC5D; SMYD2; SLIT2; TOMM20; ADORA3; PPFIA2; HMOX1; HLF; NDUFA12; ARPC1A; GNB1; PLAGL1; RAVR1; FAM102A; POU3F1; RASSF5; KCNJ3; ZNF326; TTF2; TOP1; EPS15; GGA2; ATP1A3; WDR5B; LAMB2; MICAL1; IKBKB; SCUBE1; PRKARIA; CCNE1; SAMD8; CUBN; CACYBP; CXCL16; KALRN; TCF7L1; FAM89A; SLC30A5; BCKDHB; TRIM28; GPRC5C; KTN1; GPRASP2; BTB; ARNTL; EIF2B4; CD59; CALCR; ARL6; ROCK2; PDE7B; LYPD1; KCNA4; NDUFA7; CBLN2; LDLRAP1; CC2D1A; PTPLAD1; TRAFD1; COL9A2; MAPK13; KATNA1; CSNK1A1; SCPEP1; TP53BP2; GRIA4; CD8A; TAGLN; LRRC3B; TRIM32; TBL2; PKIA; LRIG3; CAPZA2; OSTF1; ERCC8; RBM14; YTHDF1; RFXANK; COL11A1; ZNF259; FOXQ1; PRSS12; RTP1; HMGB2; RHO1; MKNK2; RRAS; ILIRAP; KPNA4; OPRL1; KCNJ6; UPF2; MOCS2; MXRA8; FADS1; BCL9; SLC12A4; STK38L; HIST1H3J; ZBTB17; AMPD2; EIF5; MRPL13; ZNF333; ATP1A1; WDR70; COG8; SPG21; CRIP2; BASP1; ZFP64; FBXO17; COL6A1; PFDN1; KIAA1033; AUP1; SLC8A3; TEX261; MAN2A1; RPL4; PRICKLE1; FBXW11; MRPL39; ANAPC4; PNRC1; PPP1R12C; BCL2L12; HAGH; LTBR; RANBP1; CDAN1; CDT1; PIK3CB; LRRC36; METAP2; GMD5; CDH26; PSME3; RBPMS2; ARIH2; ATG16L1; LRRC40; CCNB3; MAP4K4; TMEM55B; ITGA5; ARC; ZNF77; NME5; MRPS9; DUSP7; MBTPS1; NLN; CTNS; POLE3; IFT81; NSMAF; STX6; NPAS3; MIF; KIAA0586; CDK5RAP2; TP53I11; FYTDD1; PTCDD2; BST1; NXT2; RTN4IP1; BCHE; FOXJ1; CBX7; CAMK2D; ABCC9; HIC2; IPO7; PIGA; NMNAT2; ZCCHC8; FOXP2; MTCH2; TRIM2; GABRA2; RNF8; NOTCH2; MRV1; TFPT; ACSL3; FMO3; CHI3L1; CSH2; PDHB; CATSPER2; CHPT1; SMG1; SLC1A5; CD33; SOX8; TFCP2; TK2; SHQ1; ARID3A; DDX23; DPP8; VPS33A; ZNF443; COPZ1; P2RX2; YTHDF2; SYMPK; EFTUD2; IL6; LIN7B; HIPK2; CYP2A7; YIF1B; TULP3; GSTM2; P2RY2; ARNTL2; ZNF274; CD53; CD2BP2; IDI1; DOCK2; CSNK1G3; MNS1; S100A8; HMGN4; CDH23; CLIC2; AFG3L2; ELK1; STX11; MIPOL1; USP2; ADAM22; ZNF577; SYTI1; USMG5; TSGA10; FBXO16; SLC8A1; VPS54; ZDHHC16; AMICA1; CHN1; KIF1C; GPR34; TFRC; ALK; ANTXR1; AP2M1; TFEB; TM2D2; NANOG; SLC25A35; ATP2C1; PCSK6; DCTN6; ZNF234; CCNB1; GLYCTK; GLRX5; HIPK4; ARL10; ACTL6B; GPR4; ACSL6; RNF185; VCAM1; PDE6D; CLSPN; ARPC5; B3GNT6; MAT2B; FGFR4; PACSIN1; CLCN4; VPS53; AK3; PXMP4; HPCAL4; TRA2A; RNF126; PVALB; UBE2A; CDC7; RDM1; CKAP2; MSH3; EFEMP2; PCCB; CRTCL2; FBLN1; LAMA5; TRIM21; AKT3; DAND5; ATF7IP2; NKIRAS2; DCTN2; SLC9A7; NLGN1; LRRC8D; TBK1; ZNF185; MYLK; ILKAP; FSCN1; SEMA7A; EFS; LDB2; ADAM19; CYB5R2; FDPS; SLC13A3; FBXO30; LAX1; LMO7; JUND; TAOK2;

TIMP3; GPR146; RPP25; TOB1; PTMA; SULT1A1; B4GALT1; KIAA1407; ZP4; TBC1D15; GPR37; PTPN23; RNPEP; LRCH1; FILIP1; HIP1; RPS6KA5; SLC25A26; ABCC4; NAGLU; PCSK2; AQP8; EMILIN1; GRIA2; PPAP2B; ARVCF; EIF4E2; H3F3A; GOLGA4; IFT52; HLA-C; DDAH2; ADAMTS20; LZTFL1; SPAG6; GTF2B; BICD2; SPAG8; RGS9; SPSB1; TSPAN6; MED31; RPLP1; KEAP1; NPDC1; LRBA; SAP30L; SP110; TSPYL5; ANPEP; PPP3CA; YARS; ROS1; GAL3ST3; LILRA5; BST2; GCLC; GPRASP1; ACOT4; LIF; PKD1; TCERG1; RANBP2; TBC1D13; CSTB; RPS16; TDRD3; CYP2F1; MYOZ3; GSK3B; DGAT2; SLC43A3; RPL6; RAB39B; DTX2; SH3GL1; AHR; IL11; PPA1; DNASE1L2; CDH13; ZFPM2; IFIT3; NEFL; PHLDB1; CAPN5; PSMD14; CSNK1G1; NRCAM; SLC35D2; ADRBK2; GNG8; CD9; PARP16; RAB5B; ZDHHC1; POLR3H; SCGB1A1; SLC7A6; NUP43; GARS; BNIP1; TARSL2; LCN12; COL8A1; TRAPPC6B; NAPA; PPIC; SOX11; SIRT1; CSNK2A2; TMCO3; VASP; ADCY7; SLC5A11; C1QTNF6; CENPJ; CD300LF; ABL1; PROM1; FAM64A; SLC39A6; RRBPI; BID; NR4A3; DRD4; ADAM11; PKIB; TMEM87A; SLC6A16; LHFPL5; SUCLG2; NDE1; ZNF417; CLUL1; PODN; DGKZ; KCNN1; TCIRG1; ACADSB; SLC35F2; SLC6A12; CS; WDR33; OSBPL7; PAN3; WNK1; MRPS21; SLC43A2; ABCF1; TSPAN7; PUM1; DYNLT1; TRERF1; LY6H; APLP1; PAX6; NFKBIZ; NLGN3; LRRC37B; LAPTM4A; TRAF3IP1; GAB2; TECTA; TMEM9; MTX1; MORF4L1; DYRK4; GOT1; CANT1; CHSY1; IMMT; CUL2; NAB1; FLII; CXCL12; DCC; PTPN21; TMEM17; HDAC4; HAL; VPS16; SMARCC2; RBP7; DAB2; ZNF682; GOLGA7; CLTCL1; SLC16A4; DMTF1; STK35; DDX42; GTF2A2; FAM8A1; NOL6; RAB8B; RUNDC1; KCNN2; HINT3; F7; RNF165; MTMR9; HIST1H2BD; F13A1; MAPK8; HEY2; ZNF383; CECR5; EPS8L2; GLMN; NOS1AP; CYP2R1; DNAJC9; LAYN; FCRL3; POLM; GRM5; GLTP; SOX15; MYLK2; PEX7; DMPK; VPS26B; MCERS1; TRIB2; HPCAL1; TM7SF2; RNMTL1; RGR; ZMYND19; BOC; CDC16; SIPA1; TGFA; ITS2N2; GLT8D2; CLDN23; PDE2A; CNOT4; NUDT14; SLC9A8; PARP3; PPAPDC1B; POP4; PDIA4; VPS13D; NES; SERPINB1; PLCD4; FBXO44; PSME1; OTUB1; PMS2; ATP5O; PCDHB14; SQSTM1; BNIPL; SLC11A1; SNRPB; NSUN2; ZNF25; KCNK12; MRPL43; PTPLA; FBXO34; GPR62; ASCL2; SPRED2; FMO2; TPCN1; TNFRSF10D; ATP11C; ZC3HC1; GKAP1; MAGEL2; ARRB2; IL1RL1; RASGEF1A; CLK2; MYD88; SS18; CLYBL; GABPB2; KCNH1; NFATC2; SNX12; RNASEH1; GDPD4; ARL13B; QPRT; H2AFV; OTUD6B; PLEKHJ1; GUCY1B3; POLD1; BAI1; ARHGEF5; ATP2B4; ANKHD1; NTSR2; NAGS; HSD17B3; H2AFZ; ID4; RHOF; DRI; CCDC25; CYP20A1; SCIN; CALML5; APCDD1; EBI3; SCFD2; GRM2; ETFA; DDA1; UBE2H; LEPREL1; LYSMD4; SMCR7; ADH1C; NME3;

COL18A1; ANKRD40; RHOJ; ANK2; GGN; HDAC8; MAPK8IP1; INSM2; ABO; FTSJ1; TUB; KRTAP21-2; HCST; HOOK1; PRDM2; ALG8; MAPRE2; UBR2; CASC5; TBL1XR1; DDX26B; CMTM3; ALDH3B1; MLPH; SGCG; KLF6; FASTK; STAM; PDCD6IP; SPCS1; ZNF354C; NEK9; PCDHAC1; CDK2; MYCT1; SLC37A4; SLC9A3R1; MOXD1; GNB4; DGKE; RASL11B; RAB27B; SLC39A9; STAG2; LCE1E; TGFB3; STXBP1; MAPK8IP2; AGXT2L1; TIGD6; TEAD1; WDR7; GPC5; WIPI2; ACSS1; TMED5; PHF2; OPA1; PGS1; PTPRCAP; TTC8; EZH1; TRIP10; KLK6; GUCY1A2; LSM10; HS3ST1; ATXN7L2; PPM1K; CPD; FBXO36; LRRC2; UBTD1; PLP2; NDUFV3; ARHGAP26; EPB42; LHPP; OSBPL1A; ABI3BP; PPP2R5E; DNAJC19; COLEC12; MAP3K3; TRAP1; CASP7; MSX1; ITGB1BP1; SYT7; IRX3; GABRG2; NAP1L2; LMOD1; HPS4; ATP5H; USP14; CLEC2B; ABCF2; MYCN; ERBB4; SEC14L1; RBM15B; IFI30; WDFY3; CREB3L4; PIPOX; DNAH7; EPB41L5; GRPEL1; RPS29; ELOVL1; MRPS36; PTPRJ; ASGR2; NEU4; SORCS3; PLLP; CCDC15; NUDT8; KCNC4; ROPN1L; DYNC1LI2; COIL; SLC22A17; FBXL2; SPATS1; NCOR2; FAH; B3GALT2; LRRC49; KCNMB3; CAPN13; TRIM52; TRIM65; RPS27; TRIM23; ITIH4; ARSD; PON3; CMTM4; ACAD9; BOLA1; PRKCA; ARMCX2; PDCD7; CCT7; GNAO1; PRKDC; EIF4G2; ZNF222; PEX5; PRDX3; CHRDL1; RANBP6; RNF2; BDH1; RPLP2; SHANK1; CYHR1; DTNBP1; NXPH4; TSC22D1; SEMA4B; GOT2; YWHAE; HAX1; ATP6V1D; ATP8A2; ZC3H10; GCC1; NDP; GJA4; PPM1G; KCNH4; ZNF184; GNAS; CACNG1; FZD3; PPM1L; SLC6A1; SDF4; BANK1; RUNX3; SNX4; CTSS; GGT1; SSBP1; WARS; TUBGCP2; CPXM2; ZBTB5; CTBS; ARMCX1; PRAF2; PLXDC2; GNAQ; MAP3K7; NELL2; ENC1; REXO4; CASK; GALK1; FBXO22; EIF2S3; ZNF691; RAMP1; CACNG4; CPEB2; TYMS; MLC1; FAM46B; CHTF18; LAT2; ACAD10; PTPRD; DSC2; BCL2L10; PCSK4; PHF23; PHLDA1; PHKB; KLF3; TRIT1; HCN3; SETDB2; MOCS3; ACTN4; RBM23; PLA2R1; CBFA2T2; CDK2AP2; ULK2; MYO1F; TUBG2; ENPP2; SKP2; NRM; SMARCA2; ZNF514; RNF26; GABRD; ADHFE1; APEX2; MGP; NDEL1; WNT7A; HYAL2; ANKRD13C; EXOSC10; YAF2; GLUL; CLCN6; TTC14; PCNP; PRKAR2B; ZNF584; BCL10; MOV10L1; ST13; POLA2; 42068; CYP4B1; UBIAD1; RPS3; SGTB; WDR76; SLPI; FOXP4; B4GALT6; INSRR; RFC2; PAQR8; PQLC3; TREM2; CYBA; CYP3A5; NOV; NAT6; PARP8; SH3TC1; SLC25A23; NUDT16L1; PIR; PHIP; FAP; GRIN2D; FOXP1; PIK3C2G; HES6; PIGH; UPF3A; MS4A7; PARP10; USPL1; CACHD1; CYP11A1; DNHD1; TRIOBP; VPS35; CHMP4B; GRTP1; ZWINT; HLA-E; GIMAP8; RIPK1; MTF2; ZNF41; OXR1; AR; GYLTL1B; MEF2C; SLC3A1; CABIN1; IRAK3; RCBTB1; PLXDC1; KREMEN2; ATP5J2; CYB5R3; OXGR1; MTA1; ACAA2;

SLC25A24; PLA2G2D; NPHS1; ZNF423; SNRPC; CLPB; LDHA; SNX7; C3; TMEM65; CDKN1B; CACNA2D3; WIF1; PRKAG1; DCXR; VNN2; 42069; GNG4; GPR37L1; GRM3; SYN2; CREB3; NMI; CYP2C18; TACC3; GTF3C5; FAM63A; GRHL1; TAL1; TNIP1; CRELD1; ZNF675; TNNT3; ZNF646; GLO1; CHMP6; DNAJC1; CRELD2; NAV1; CASKIN2; PLXNC1; NOTCH4; ADPGK; UCK1; OPA3; MTPP; FBN1; WDR54; RBMX; SELK; FATE1; CCT4; PRPSAP1; AGPAT2; CRISPLD2; WNT10B; CHD2; CLN6; BTG4; ODF1; GRIN2C; GIMAP7; SEMA6A; UBE2M; RAB33B; CYP1A1; UNC93B1; TTLL1; CDIPT; LETM1; FARP2; IL18; CORO1B; POLR2F; GALK2; SH2D3C; AXL; AK1; FUT3; XPA; CEL; MSR1; CASQ1; NDUFA5; TMEM60; SORL1; OSCAR; PTGES; BLOC1S2; SYNJ1; NRXN2; ATP6V1H; DHFRL1; NUP35; CORO6; AKAP7; ARV1; GFM2; RELA; NGFR; NIPA2; GLRX2; and TSPAN5.

[00117] In some embodiments, the genes whose expression are significantly correlated with chronological age are determined using a false discovery rate of $< 5\%$ by linear regression, and after correction for gender and batch effects.

[00118] In some embodiments, the genes whose expression are significantly correlated with chronological age are determined using microarray datasets from one or more cohorts or plurality of individuals. In some embodiments, the microarray datasets are transcriptome-wide microarray datasets. In some embodiments, the microarray datasets are available in the art, including, but not limited to, Tgen (Myers et al., 2007; Webster et al., 2009), BrainEqtl (Gibbs et al., 2010), HBTRC (Zhang et al., 2013), and BrainCloud (Colantuoni et al., 2011).

[00119] *Selecting genes associated with aging*

[00120] In certain embodiments, the invention provides methods for identifying and selecting genes correlated with chronological age. Chronological age refers to the actual age of the individual at the time the sample from which the gene expression levels are determined was taken.

[00121] In a given cohort (plurality of individuals), individual genes are identified which exhibit expression levels that positively or negatively correlate with chronological age. Linear regression can be used to characterize the correlation between chronological age and gene expression levels. For example, the linear correlation between chronological age and

gene expression levels is assessed using R's *lm()* function. Gene expression is plotted as a function of chronological age for every individual in a given cohort for each gene. In some embodiments, the genes whose expression is significantly correlated with chronological age are the genes p-values of less than 0.05, less than 0.01, or less than 0.001. In some embodiments, the false discovery cut-off threshold for genes significantly correlated with chronological age is 1%, 5%, or 10%.

[00122] In certain embodiments, cohorts (plurality of individuals) used in the methods described herein are subdivided into selected age groups. For example, a late-life cohort can be studied independently from other age-based cohorts. In some embodiments, the cohort includes individual who are over 25 years old. In some embodiments, the cohort includes individual who are 25 years old and under. In some embodiments, the cohort includes individual who are over 40 years old. In some embodiments, the cohort of includes individual who are 45 years old and under. In some embodiments, the cohort includes individual who are over 45 years old. In some embodiments, the cohort of includes individual who are 45 years old and under. In some embodiments, the cohort includes individual who are over 50 years old. In some embodiments, the cohort of includes individual who are 55 years old and under. In some embodiments, the cohort includes individual who are over 60 years old. In some embodiments, the cohort of includes individual who are 60 years old and under. In some embodiments, the cohort includes individual who are over 65 years old. In some embodiments, the cohort of includes individual who are 65 years old and under. In some embodiments, the cohort includes individual who are over 70 years old. In some embodiments, the cohort of includes individual who are 75 years old and under. In some embodiments, the cohort includes individual who are over 80 years old. In some embodiments, the cohort includes individual who are 80 years old and under. In some embodiments, the cohort includes individual who are over 85 years old. In some embodiments, the cohort of includes individual who are 85 years old and under. In some embodiments, the cohort includes individual who are over 90 years old. In some embodiments, the cohort of includes individual who are 90 years old and under. In some embodiments, the cohort includes individual who are over 95 years old. In some embodiments, the cohort of includes individual who are 95 years old and under. In some embodiments, the cohort includes individuals who are between 60 and 65 years old. In some embodiments, the cohort includes individuals who are between 65 and 80 years old. In some embodiments, the cohort includes individuals who are between 80 and 90 years old.

[00123] In certain embodiments, cohorts (plurality of individuals) used in the methods described herein are subdivided into selected disease status. In some embodiments, the cohort comprises individuals with or at risk of developing Alzheimer's Disease. In some embodiments, the cohort comprises individuals with or at risk of developing Amyotrophic Lateral Sclerosis. In some embodiments, the cohort comprises individuals with or at risk of developing Hippocampal Sclerosis. In some embodiments, the cohort comprises individuals with or at risk of developing Parkinson's Disease. In some embodiments, the cohort comprises individuals with or at risk of developing Fronto-temporal dementia. In some embodiments, the cohort comprises individuals that are healthy.

[00124] In certain embodiments, disease status of an individual is determined by any suitable method, including but not limited to a physical examination of the subject, a neurological examination of the subject, a brain scan, or a combination thereof. Suitable methods for determining the disease status are those known to one of skill in the art. In some embodiments, the subject is not diagnosed with any disease. In another embodiment, the subject is diagnosed with a disease. In another embodiment, the subject is diagnosed with a pre-disease state.

[00125] In one embodiment, the method further comprises a physical examination of the subject, a neurological examination of the subject, a brain scan, or a combination thereof. Methods and types of physical examinations are known to one of skill in the art.

[00126] *Determining Δ -aging*

[00127] In certain embodiments, the invention provides methods for determining differential-aging (Δ -aging). Differential-aging is defined as the difference between predicted biological age (based on the aggregate of expression levels of age-dependent transcripts) and chronological age for each individual within a given cohort. Differential-aging is expressed as a numerical value in time units.

[00128] For each given gene, the gene specific differential-aging value is the difference between the apparent biological age, imputed based on the expression level of the gene, and the actual chronological age. For example, gene expression as a function of chronological age is plotted for a given cohort of individuals, and linear regression is used to generate a regression line of a gene's expression levels as a function of chronological age across the entire cohort. The differential-age for an individual for a given gene is expressed as the ratio

between the residual value (how much the expression of the given gene deviates from the regression line for that particular gene) and the coefficient obtained by linear regression of the expression level of the given gene as a function of chronological age across the entire cohort. The ratio for each gene significantly correlated with chronological age in the cohort, are aggregated by integration to provide the biological age of the sample of the subject. Accordingly, the biological age represents the aggregate expression levels of the age-dependent transcripts in a sample from an individual. The difference between the biological age and the actual chronological age of the individual corresponded to the differential aging or Δ -aging trait.

[00129] In certain embodiments, cohorts (plurality of individuals) used in the methods described herein are subdivided into selected age groups. For example, a late-life cohort can be studied independently from other age-based cohorts. In some embodiments, the cohort includes individual who are over 25 years old. In some embodiments, the cohort includes individual who are 25 years old and under. In some embodiments, the cohort includes individual who are over 40 years old. In some embodiments, the cohort of includes individual who are 45 years old and under. In some embodiments, the cohort includes individual who are over 45 years old. In some embodiments, the cohort of includes individual who are 45 years old and under. In some embodiments, the cohort includes individual who are over 50 years old. In some embodiments, the cohort of includes individual who are 55 years old and under. In some embodiments, the cohort includes individual who are over 60 years old. In some embodiments, the cohort of includes individual who are 60 years old and under. In some embodiments, the cohort includes individual who are over 65 years old. In some embodiments, the cohort of includes individual who are 65 years old and under. In some embodiments, the cohort includes individual who are over 70 years old. In some embodiments, the cohort of includes individual who are 75 years old and under. In some embodiments, the cohort includes individual who are over 80 years old. In some embodiments, the cohort includes individual who are 80 years old and under. In some embodiments, the cohort includes individual who are over 85 years old. In some embodiments, the cohort of includes individual who are 85 years old and under. In some embodiments, the cohort includes individual who are over 90 years old. In some embodiments, the cohort of includes individual who are 90 years old and under. In some embodiments, the cohort includes individual who are over 95 years old. In some embodiments, the cohort of includes individual who are 95 years old and under. In

some embodiments, the cohort includes individuals who are between 60 and 65 years old. In some embodiments, the cohort includes individuals who are between 65 and 80 years old. In some embodiments, the cohort includes individuals who are between 80 and 90 years old.

[00130] In certain embodiments, cohorts (plurality of individuals) used in the methods described herein are subdivided into selected disease status. In some embodiments, the cohort comprises individuals with or at risk of developing Alzheimer's Disease. In some embodiments, the cohort comprises individuals with or at risk of developing Amyotrophic Lateral Sclerosis. In some embodiments, the cohort comprises individuals with or at risk of developing Hippocampal Sclerosis. In some embodiments, the cohort comprises individuals with or at risk of developing Parkinson's Disease. In some embodiments, the cohort comprises individuals with or at risk of developing Fronto-temporal dementia. In some embodiments, the cohort comprises individuals that are healthy.

[00131] In certain embodiments, disease status of an individual is determined by any suitable method, including but not limited to a physical examination of the subject, a neurological examination of the subject, a brain scan, or a combination thereof. Suitable methods for determining the disease status are those known to one of skill in the art. In some embodiments, the subject is not diagnosed with any disease. In another embodiment, the subject is diagnosed with a disease. In another embodiment, the subject is diagnosed with a pre-disease state.

[00132] In one embodiment, the method further comprises a physical examination of the subject, a neurological examination of the subject, a brain scan, or a combination thereof. Methods and types of physical examinations are known to one of skill in the art.

[00133] *Modifiers of Δ -aging*

[00134] In certain embodiments, the invention provides methods for determining genetic modifiers of biological aging using differential-aging as a quantitative trait. For example, genetic modifiers and non-genetic modifiers, such as, but not limited to environmental factors and anti-aging interventions (e.g. exercise, diet, lifestyle) are identified. GWAS can be used to identify SNPs and genes with strong associations with the differential-aging trait.

[00135] As described herein, the methods of the invention provide improvements over prior technology. Prior research has focused on identifying and selecting genes with clear age-

associated phenotypes. The present invention is an improvement over current methods of determining differential-aging which identifies age-associated traits in healthy individuals. By identifying genes associated with differential-aging across the population and genetic modifiers, age-associated disorders can be treated with genetic modulators. The method described here provides an improved approach because it provides a quantitative trait, Δ -aging, that allows for the identification of genetic modifiers of the trait. The methods described herein are unbiased and use no prior assumption on the nature of age-associated phenotypic changes. Prior studies have been used to identify age-associated phenotypes but most are likely secondary to aging and not causal.

[00136] *Computer program products for determining biological age and differential aging*

[00137] In certain aspects, the invention provides a computer program product for determining a biological age of a sample from a subject comprising instructions which, when the program is executed by a computer, cause the computer to carry out the steps of: a) providing a gene expression level of a plurality of genes in a sample for a plurality of individuals; b) performing a linear regression of the expression level of each gene of the plurality of genes as a function of chronological age of each of the plurality of individuals to determine a linear regression co-efficient for each gene of the plurality of genes; c) selecting, from the plurality of genes, the genes whose expression level is significantly correlated with chronological age; d) determining, for each gene whose expression level is significantly correlated with chronological age selected in step (c), a residual value, wherein the residual value is determined by comparing the gene expression level of said gene in the sample from the subject to the linear regression performed in step (b) for said gene; e) determining, for each gene whose expression level is significantly correlated with chronological age, a ratio of the residual determined in step (d) to the linear regression coefficient determined in step (b); f) repeating steps (d) and (e) for each gene whose expression level is significantly correlated with chronological age selected in step (c); and g) integrating the ratios for each gene whose expression level is significantly correlated with chronological age, wherein the integrated ratio corresponds to the biological age of the sample from the subject.

[00138] In some embodiments, the computer program product further comprises carrying out the step of: comparing the biological age to a chronological age of the subject to determine a differential aging value for the subject. In some embodiments, the plurality of

individuals includes the subject. In some embodiments, wherein steps (d) to (g) are performed for each individual in the plurality of individuals.

[00139] In some embodiments, the computer program product further comprises carrying out the step of performing a genome-wide association study (GWAS). In some embodiments, the GWAS identifies single-nucleotide polymorphism (SNP) modifiers of the differential aging value in the plurality of individuals. In some embodiments, the GWAS identifies genetic modifiers of the differential aging value in the plurality of individuals. In some embodiments, the gene expression level is the gene expression level in the brain. In some embodiments, the gene expression level is the gene expression level in the frontal cortex. In some embodiments, wherein the gene expression level is the gene expression level in the cerebellum.

[00140] In certain aspects, the invention provides an apparatus configured to determine a biological age of a sample from a subject. In certain aspects, the invention provides an apparatus configured to determine differential aging of a sample from a subject. In some embodiments, the apparatus is configured to a) provide a gene expression level of a plurality of genes in a sample for a plurality of individuals; b) perform a linear regression of the expression level of each gene of the plurality of genes as a function of chronological age of each of the plurality of individuals to determine a linear regression co-efficient for each gene of the plurality of genes; c) select, from the plurality of genes, the genes whose expression level is significantly correlated with chronological age; d) determine, for each gene whose expression level is significantly correlated with chronological age selected in step (c), a residual value, wherein the residual value is determined by comparing the gene expression level of said gene in the sample from the subject to the linear regression performed in step (b) for said gene; e) determine, for each gene whose expression level is significantly correlated with chronological age, a ratio of the residual determined in step (d) to the linear regression coefficient determined in step (b); f) repeat steps (d) and (e) for each gene whose expression level is significantly correlated with chronological age selected in step (c); and g) integrate the ratios for each gene whose expression level is significantly correlated with chronological age, wherein the integrated ratio corresponds to the biological age of the sample from the subject.

[00141] In some embodiments, the apparatus further comprises carrying out the step of: comparing the biological age to a chronological age of the subject to determine a differential

aging value for the subject. In some embodiments, the plurality of individuals includes the subject. In some embodiments, wherein steps (d) to (g) are performed for each individual in the plurality of individuals.

[00142] In some embodiments, the apparatus further comprises carrying out the step of performing a genome-wide association study (GWAS). In some embodiments, the GWAS identifies single-nucleotide polymorphism (SNP) modifiers of the differential aging value in the plurality of individuals. In some embodiments, the GWAS identifies genetic modifiers of the differential aging value in the plurality of individuals. In some embodiments, the gene expression level is the gene expression level in the brain. In some embodiments, the gene expression level is the gene expression level in the frontal cortex. In some embodiments, wherein the gene expression level is the gene expression level in the cerebellum.

[00143] In certain aspects, the invention provides a system comprising an apparatus configured to determine a biological age of a sample from a subject. In certain aspects, the invention provides a system comprising an apparatus configured to determine differential aging of a sample from a subject. In some embodiments, the system comprises an apparatus configured to a) provide a gene expression level of a plurality of genes in a sample for a plurality of individuals; b) perform a linear regression of the expression level of each gene of the plurality of genes as a function of chronological age of each of the plurality of individuals to determine a linear regression co-efficient for each gene of the plurality of genes c) select, from the plurality of genes, the genes whose expression level is significantly correlated with chronological age; d) determine, for each gene whose expression level is significantly correlated with chronological age selected in step (c), a residual value, wherein the residual value is determined by comparing the gene expression level of said gene in the sample from the subject to the linear regression performed in step (b) for said gene; e) determining, for each gene whose expression level is significantly correlated with chronological age, a ratio of the residual determined in step (d) to the linear regression coefficient determined in step (b); f) repeat steps (d) and (e) for each gene whose expression level is significantly correlated with chronological age selected in step (c); and g) integrate the ratios for each gene whose expression level is significantly correlated with chronological age, wherein the integrated ratio corresponds to the biological age of the sample from the subject. In some embodiments, the system further comprises an apparatus configured to compare the biological age to a chronological age of the subject to determine a differential aging value for the subject.

[00144] In some embodiments, the system further comprises carrying out the step of: comparing the biological age to a chronological age of the subject to determine a differential aging value for the subject. In some embodiments, the plurality of individuals includes the subject. In some embodiments, wherein steps (d) to (g) are performed for each individual in the plurality of individuals.

[00145] In some embodiments, the system further comprises carrying out the step of performing a genome-wide association study (GWAS). In some embodiments, the GWAS identifies single-nucleotide polymorphism (SNP) modifiers of the differential aging value in the plurality of individuals. In some embodiments, the GWAS identifies genetic modifiers of the differential aging value in the plurality of individuals. In some embodiments, the gene expression level is the gene expression level in the brain. In some embodiments, the gene expression level is the gene expression level in the frontal cortex. In some embodiments, wherein the gene expression level is the gene expression level in the cerebellum.

[00146] In certain aspects, the invention provides a computer program product for determining a biological age of a sample from a subject comprising instructions which, when the program is executed by a computer, cause the computer to carry out the steps of: a) providing a gene expression level of a plurality of genes in a sample for a plurality of individuals; b) determining, for each gene whose expression level is significantly correlated with chronological age, a residual value, wherein the residual value is determined by comparing the gene expression level of said gene in the sample from the subject to a linear regression of the expression level of the gene as a function of chronological age for a plurality of individuals; c) determining, for each gene whose expression level is significantly correlated with chronological age, a ratio of the residual value determined in step (b) to the linear regression coefficient for the gene; d) repeating steps (b) and (c) for each gene whose expression level is significantly correlated with chronological age; and e) integrating the ratios for each gene whose expression level is significantly correlated with chronological age, wherein the integrated ratio corresponds to the biological age of the sample from the subject.

[00147] In some embodiments, the computer program product further comprises carrying out the step of: comparing the biological age to a chronological age of the subject to determine a differential aging value for the subject. In some embodiments, the plurality of individuals includes the subject. In some embodiments, steps (c) to (e) are performed for each individual in the plurality of individuals.

[00148] In some embodiments, the computer program product further comprises carrying out the step of performing a genome-wide association study (GWAS). In some embodiments, the GWAS identifies single-nucleotide polymorphism (SNP) modifiers of the differential aging value in the plurality of individuals. In some embodiments, the GWAS identifies genetic modifiers of the differential aging value in the plurality of individuals. In some embodiments, the gene expression level is the gene expression level in the brain. In some embodiments, the gene expression level is the gene expression level in the frontal cortex. In some embodiments, wherein the gene expression level is the gene expression level in the cerebellum.

[00149] In certain aspects, the invention provides an apparatus configured to determine a biological age of a sample from a subject. In certain aspects, the invention provides an apparatus configured to determine differential aging of a sample from a subject. In some embodiments, the apparatus is configured to a) provide a gene expression level of a plurality of genes in a sample for a plurality of individuals; b) determine, for each gene whose expression level is significantly correlated with chronological age, a residual value, wherein the residual value is determined by comparing the gene expression level of said gene in the sample from the subject to a linear regression of the expression level of the gene as a function of chronological age for a plurality of individuals; c) determine, for each gene whose expression level is significantly correlated with chronological age, a ratio of the residual value determined in step (b) to the linear regression coefficient for the gene; d) repeat steps (b) and (c) for each gene whose expression level is significantly correlated with chronological age; and e) integrate the ratios for each gene whose expression level is significantly correlated with chronological age, wherein the integrated ratio corresponds to the biological age of the sample from the subject. In some embodiments, the system further comprises an apparatus configured to compare the biological age to a chronological age of the subject to determine a differential aging value for the subject

[00150] In some embodiments, the apparatus further comprises carrying out the step of: comparing the biological age to a chronological age of the subject to determine a differential aging value for the subject. In some embodiments, the plurality of individuals includes the subject. In some embodiments, steps (c) to (e) are performed for each individual in the plurality of individuals.

[00151] In some embodiments, the apparatus further comprises carrying out the step of performing a genome-wide association study (GWAS). In some embodiments, the GWAS identifies single-nucleotide polymorphism (SNP) modifiers of the differential aging value in the plurality of individuals. In some embodiments, the GWAS identifies genetic modifiers of the differential aging value in the plurality of individuals. In some embodiments, the gene expression level is the gene expression level in the brain. In some embodiments, the gene expression level is the gene expression level in the frontal cortex. In some embodiments, wherein the gene expression level is the gene expression level in the cerebellum.

[00152] In certain aspects, the invention provides a system comprising an apparatus configured to determine a biological age of a sample from a subject. In certain aspects, the invention provides a system comprising an apparatus configured to determine differential aging of a sample from a subject. In some embodiments, the system comprises an apparatus configured to a) provide a gene expression level of a plurality of genes in a sample for a plurality of individuals; b) determine, for each gene whose expression level is significantly correlated with chronological age, a residual value, wherein the residual value is determined by comparing the gene expression level of said gene in the sample from the subject to a linear regression of the expression level of the gene as a function of chronological age for a plurality of individuals; c) determine, for each gene whose expression level is significantly correlated with chronological age, a ratio of the residual value determined in step (b) to the linear regression coefficient for the gene; d) repeat steps (b) and (c) for each gene whose expression level is significantly correlated with chronological age; and e) integrate the ratios for each gene whose expression level is significantly correlated with chronological age, wherein the integrated ratio corresponds to the biological age of the sample from the subject. In some embodiments, the system further comprises an apparatus configured to compare the biological age to a chronological age of the subject to determine a differential aging value for the subject.

[00153] In some embodiments, the system further comprises carrying out the step of: comparing the biological age to a chronological age of the subject to determine a differential aging value for the subject. In some embodiments, the plurality of individuals includes the subject. In some embodiments, steps (c) to (e) are performed for each individual in the plurality of individuals.

[00154] In some embodiments, the system further comprises carrying out the step of performing a genome-wide association study (GWAS). In some embodiments, the GWAS identifies single-nucleotide polymorphism (SNP) modifiers of the differential aging value in the plurality of individuals. In some embodiments, the GWAS identifies genetic modifiers of the differential aging value in the plurality of individuals. In some embodiments, the gene expression level is the gene expression level in the brain. In some embodiments, the gene expression level is the gene expression level in the frontal cortex. In some embodiments, wherein the gene expression level is the gene expression level in the cerebellum.

[00155] *Methods of determining genetic modifiers of a phenotype associated with a genotype of interest*

[00156] Genome-wide association studies are used to probe the association between many genotypes and a single variable of interest (Van Deerlin VM, et al. Common variants at 7p21 are associated with frontotemporal lobar degeneration with TDP-43 inclusions. Nat Genet. 2010 Mar; 42(3): pp. 234-9). While specific, causal mutations have been identified for many complex diseases, variants in other genes can act as genetic modifiers, also affecting the presentation and severity of the associated phenotypes (Gallagher MD, et al. TMEM106B is a genetic modifier of frontotemporal lobar degeneration with C9orf72 hexanucleotide repeat expansions. Acta Neuropathol. 2014 Mar; 127(3): pp. 407-18.). Moreover, causal genes are often not feasible direct pharmacological targets, and as such, there is a need for use of transcriptome-wide gene expression analysis to systematically identify targetable genes and pathways (Caberlotto L, et al. Integration of transcriptomic and genomic data suggests candidate mechanisms for APOE4-mediated pathogenic action in Alzheimer's disease. Sci Rep. 2016 Sep; 6:32583).

[00157] Described herein is the use of transcriptomic techniques and approaches to discover new druggable targets and pathways of interest for complex diseases. Genes of interest include those that mimic, potentiate, or ameliorate the phenotypes associated with variants of known causative genes. Through the analysis of genotypes and other factors that affect gene expression levels of various gene sets in many individuals, the invention also allows for the determination of the pathways that are the most modified by genetic variants of interest. The reversion of the coordinated variation in pathway gene expression to that of unaffected individuals can provide a therapeutic option. For example, the invention has been used identify ILR2A as a genetic modifier and synaptogenesis as a pathway of interest for

individuals carrying the TMEM106B risk allele, which is associated with increased synaptic loss in neurodegenerative disease.

[00158] In certain embodiments, the invention provides methods for determining delta-genotype of interest (Δ -genotype of interest). Delta-genotype of interest corresponds to the difference between the aggregate of expression levels of each gene in a plurality of genes whose expression is correlated with the genotype of interest for an individual and the aggregate of expression levels of each gene of the plurality of genes for a given cohort of individuals with the known genotype of interest. For example, the expression level of a given set of genes may be correlated with a particular genotype (i.e. a genotype of interest has a phenotype associated with it, wherein the phenotype corresponds to the expression level of a particular set of genes). The aggregate expression levels of this set of genes can then be determined as a quantitative trait that represents how much the phenotype of a sample is similar or different to the phenotype associated with the genotype of interest. The genetic modifiers of the quantitative trait can then be determined.

[00159] For each given gene, the gene specific delta-genotype of interest value is the difference between gene expression level in a sample and regression line for that gene. For example, gene expression as a function of a genotype of interest is plotted for a given cohort of individuals, and linear regression is used to generate a regression line of a gene's expression level as a function of the genotype of interest across the entire cohort. The delta-genotype of interest for an individual for a given gene is expressed as the ratio between the residual value (how much the expression of the given gene deviates from the regression line for that particular gene) and the coefficient obtained by linear regression of the expression level of the given gene as a function of the genotype of interest across the entire cohort. The ratio for each gene significantly correlated with the genotype of interest in the cohort, are aggregated by integration. Accordingly, this quantitative trait represents the aggregate expression levels of the genotype-dependent transcripts in a sample from an individual.

[00160] In certain aspects, the invention provides a computer-implemented method of identifying one or more genetic modifiers of a phenotype associated with a genotype of interest comprising: a) providing a gene expression level of a plurality of genes in a sample for a plurality of individuals; b) performing a linear regression of the expression level of each gene of the plurality of genes as a function of the genotype of interest of each of the plurality of individuals to determine a linear regression co-efficient for each gene of the plurality of

genes; c) selecting, from the plurality of genes, the genes significantly associated with the genotype of interest, wherein the genes selected and their expression levels corresponds to the phenotype associated with the genotype of interest; d) determining a residual value for each gene significantly associated with genotype of interest by comparing the gene expression level of said gene in a sample from a subject to the linear regression for said gene; e) determining a ratio of the residual value to the linear regression coefficient for each gene significantly associated with the genotype of interest; f) integrating the ratios for each gene significantly associated with the genotype of interest to determine a phenotype of the sample from the subject; g) comparing the phenotype of the sample to the phenotype associated with the genotype of interest to determine a differential genotype of interest value for the subject; h) performing steps (d) to (g) for each individual in the plurality of individuals; and i) identifying genetic modifiers that modulate the differential genotype of interest value.

[00161] In some embodiments, the method further comprises performing a genome-wide association study (GWAS). In some embodiments, the plurality of individuals includes the subject. In some embodiments, the GWAS identifies single-nucleotide polymorphism (SNP) modifiers of the differential genotype of interest value in the plurality of individuals. In some embodiments, the genetic modifier is a single-nucleotide polymorphism (SNP). In some embodiments, the genotype of interest is the risk allele of TMEM106B. In some embodiments, the risk allele of TMEM106B is and A at SNP rs1990622. In some embodiments, the genotype of interest is a risk allele associated with a disease or disorder. In some embodiments, the gene expression level is the gene expression level in the brain. In some embodiments, the gene expression level is the gene expression level in the frontal cortex. In some embodiments, the gene expression level is the gene expression level in the cerebellum.

[00162] In certain aspects, the invention provides a computer-implemented method of determining a phenotype of a sample from a subject, wherein the phenotype is correlated with a haplotype of interest, the method comprising: a) providing a gene expression level of a plurality of genes in a sample for a plurality of individuals, wherein each individual's haplotype of interest is known; b) performing a linear regression of the expression level of each gene of the plurality of genes as a function of the haplotype of interest for each of the plurality of individuals to determine a linear regression co-efficient for each gene of the plurality of genes; c) selecting, from the plurality of genes, the genes whose expression level

is significantly correlated with the haplotype of interest; d) determining, for each gene whose expression level is significantly correlated with the haplotype of interest selected in step (c), a residual value, wherein the residual value is determined by comparing the gene expression level of said gene in a sample from a subject to the linear regression performed in step (b) for said gene; e) determining, for each gene whose expression level is significantly correlated with the haplotype of interest, a ratio of the residual value determined in step (d) to the linear regression coefficient determined in step (b); f) repeating steps (d) and (e) for each gene whose expression level is significantly correlated with the haplotype of interest selected in step (c); and g) integrating the ratios for each gene whose expression level is significantly correlated with the haplotype of interest, wherein the integrated ratio corresponds to the phenotype of the sample from the subject.

[00163] In some embodiments, the plurality of individuals includes the subject. In some embodiments, steps (d) to (g) are performed for each individual in the plurality of individuals.

[00164] In some embodiments, the method further comprises: h) performing a genome-wide association study (GWAS). In some embodiments, the GWAS identifies single-nucleotide polymorphism (SNP) modifiers of the phenotype correlated with the haplotype of interest in the plurality of individuals. In some embodiments, the GWAS identifies genetic modifiers of the phenotype in the plurality of individuals. In some embodiments, the haplotype of interest is defined as 0, 1, or 2 allele copies. In some embodiments, the allele copies are determined by SNP genotyping. In some embodiments, the phenotype correlated with the haplotype of interest is an expression level of a plurality of genes.

[00165] In some embodiments, the gene expression level is the gene expression level in the brain. In some embodiments, the gene expression level is the gene expression level in the frontal cortex. In some embodiments, the gene expression level is the gene expression level in the cerebellum.

[00166] In certain aspects, the invention provides a computer-implemented method of identifying one or more genetic modifiers of a plurality of genes whose expression level is correlated with a genotype of interest comprising: a) providing a gene expression level of a plurality of genes in a sample for a plurality of individuals, wherein each individual's genotype of interest is known; b) performing a linear regression of the expression level of each gene of the plurality of genes as a function of the genotype of interest of each of the

plurality of individuals to determine a linear regression co-efficient for each gene of the plurality of genes; c) selecting, from the plurality of genes, the genes whose expression level is significantly correlated with the genotype of interest; d) determining, for each gene whose expression level is significantly correlated with the genotype of interest selected in step (c), a residual value, wherein the residual value is determined by comparing the gene expression level of said gene in a sample from a subject to the linear regression performed in step (b) for said gene; e) determining, for each gene whose expression level is significantly correlated with the genotype of interest, a ratio of the residual value determined in step (d) to the linear regression coefficient determined in step (b); f) repeating steps (d) and (e) for each gene whose expression level is significantly correlated with the genotype of interest selected in step (c); g) integrating the ratios for each gene significantly correlated with the genotype of interest; h) performing steps (d) to (g) for each individual in the plurality of individuals; and j) identifying genetic modifiers that modulate the integrated ratio value determined in step (g).

[00167] In some embodiments, the genetic modifiers that modulate the integrated ratio value determined in step (g) are determined by performing a genome-wide association study (GWAS). In some embodiments, the plurality of individuals includes the subject.

[00168] In some embodiments, the GWAS identifies single-nucleotide polymorphism (SNP) modifiers of the integrated ratio value determined in step (g) in the plurality of individuals. In some embodiments, the genetic modifier is a single-nucleotide polymorphism (SNP). In some embodiments, the genotype of interest is the risk allele of TMEM106B. In some embodiments, the risk allele of TMEM106B is and A at SNP rs1990622. In some embodiments, the genotype of interest is a risk allele associated with a disease or disorder. In some embodiments, the genotype of interest is a non-risk allele. In some embodiments, the genotype of interest is a haplotype.

[00169] In some embodiments, the gene expression level is the gene expression level in the brain. In some embodiments, the gene expression level is the gene expression level in the frontal cortex. In some embodiments, the gene expression level is the gene expression level in the cerebellum.

[00170] In some embodiments, the plurality of individuals are healthy individuals. In some embodiments, the plurality of individuals have a disease or disorder. In some embodiments,

the plurality of individuals have a neurodegenerative disease. In some embodiments, the neurodegenerative disease is Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, or Fronto-temporal dementia.

[00171] In certain embodiments, cohorts (plurality of individuals) used in the methods described herein are subdivided into selected age groups. For example, a late-life cohort can be studied independently from other age-based cohorts. In some embodiments, the cohort includes individual who are over 25 years old. In some embodiments, the cohort includes individual who are 25 years old and under. In some embodiments, the cohort includes individual who are over 40 years old. In some embodiments, the cohort includes individual who are 40 years old and under. In some embodiments, the cohort includes individual who are over 45 years old. In some embodiments, the cohort includes individual who are 45 years old and under. In some embodiments, the cohort includes individual who are over 50 years old. In some embodiments, the cohort includes individual who are 50 years old and under. In some embodiments, the cohort includes individual who are over 55 years old. In some embodiments, the cohort includes individual who are 55 years old and under. In some embodiments, the cohort includes individual who are over 60 years old. In some embodiments, the cohort includes individual who are 60 years old and under. In some embodiments, the cohort includes individual who are over 65 years old. In some embodiments, the cohort includes individual who are 65 years old and under. In some embodiments, the cohort includes individual who are over 70 years old. In some embodiments, the cohort includes individual who are 70 years old and under. In some embodiments, the cohort includes individual who are over 75 years old. In some embodiments, the cohort includes individual who are 75 years old and under. In some embodiments, the cohort includes individual who are over 80 years old. In some embodiments, the cohort includes individual who are 80 years old and under. In some embodiments, the cohort includes individual who are over 85 years old. In some embodiments, the cohort includes individual who are 85 years old and under. In some embodiments, the cohort includes individual who are over 90 years old. In some embodiments, the cohort includes individual who are 90 years old and under. In some embodiments, the cohort includes individual who are over 95 years old. In some embodiments, the cohort includes individual who are 95 years old and under. In some embodiments, the cohort includes individuals who are between 60 and 65 years old. In some embodiments, the cohort includes individuals who are between 65 and 80 years old. In some embodiments, the cohort includes individuals who are between 80 and 90 years old.

[00172] In certain embodiments, cohorts (plurality of individuals) used in the methods described herein are subdivided into selected disease status. In some embodiments, the

cohort comprises individuals with or at risk of developing Alzheimer's Disease. In some embodiments, the cohort comprises individuals with or at risk of developing Amyotrophic Lateral Sclerosis. In some embodiments, the cohort comprises individuals with or at risk of developing Hippocampal Sclerosis. In some embodiments, the cohort comprises individuals with or at risk of developing Parkinson's Disease. In some embodiments, the cohort comprises individuals with or at risk of developing Fronto-temporal dementia. In some embodiments, the cohort comprises individuals that are healthy.

[00173] In certain embodiments, disease status of an individual is determined by any suitable method, including but not limited to a physical examination of the subject, a neurological examination of the subject, a brain scan, or a combination thereof. Suitable methods for determining the disease status are those known to one of skill in the art. In some embodiments, the subject is not diagnosed with any disease. In another embodiment, the subject is diagnosed with a disease. In another embodiment, the subject is diagnosed with a pre-disease state.

[00174] In one embodiment, the method further comprises a physical examination of the subject, a neurological examination of the subject, a brain scan, or a combination thereof. Methods and types of physical examinations are known to one of skill in the art.

[00175] *Computer program products for determining genetic modifiers of a phenotype associated with a genotype of interest*

[00176] In certain aspects, the invention provides a computer program product for identifying one or more genetic modifiers of a phenotype associated with a genotype of interest comprising instructions which, when the program is executed by a computer, cause the computer to carry out the steps of: a) providing a gene expression level of a plurality of genes in a sample for a plurality of individuals; b) performing a linear regression of the expression level of each gene of the plurality of genes as a function of the genotype of interest of each of the plurality of individuals to determine a linear regression co-efficient for each gene of the plurality of genes; c) selecting, from the plurality of genes, the genes significantly associated with the genotype of interest, wherein the genes selected and their expression levels corresponds to the phenotype associated with the genotype of interest; d) determining a residual value for each gene significantly associated with genotype of interest by comparing the gene expression level of said gene in a sample from a subject to the linear

regression for said gene; e) determining a ratio of the residual value to the linear regression coefficient for each gene significantly associated with the genotype of interest; f) integrating the ratios for each gene significantly associated with the genotype of interest to determine a phenotype of the sample from the subject; g) comparing the phenotype of the sample to the phenotype associated with the genotype of interest to determine a differential genotype of interest value for the subject; h) performing steps (d) to (g) for each individual in the plurality of individuals; and i) identifying genetic modifiers that modulate the differential genotype of interest value.

[00177] In certain aspects, the invention provides a computer program product for determining the phenotype of a sample from a subject, wherein the phenotype is correlated with a haplotype of interest, comprising instructions which, when the program is executed by a computer, cause the computer to carry out the steps of: a) providing a gene expression level of a plurality of genes in a sample for a plurality of individuals, wherein each individual's haplotype of interest is known; b) performing a linear regression of the expression level of each gene of the plurality of genes as a function of the haplotype of interest for each of the plurality of individuals to determine a linear regression co-efficient for each gene of the plurality of genes; c) selecting, from the plurality of genes, the genes whose expression level is significantly correlated with the haplotype of interest; d) determining, for each gene whose expression level is significantly correlated with the haplotype of interest selected in step (c), a residual value, wherein the residual value is determined by comparing the gene expression level of said gene in a sample from a subject to the linear regression performed in step (b) for said gene; e) determining, for each gene whose expression level is significantly correlated with the haplotype of interest, a ratio of the residual value determined in step (d) to the linear regression coefficient determined in step (b); f) repeating steps (d) and (e) for each gene whose expression level is significantly correlated with the haplotype of interest selected in step (c); and g) integrating the ratios for each gene whose expression level is significantly correlated with the haplotype of interest, wherein the integrated ratio corresponds to the phenotype of the sample from the subject.

[00178] In some embodiments, the plurality of individuals includes the subject. In some embodiments, steps (d) to (g) are performed for each individual in the plurality of individuals.

[00179] In some embodiments, the computer program product further comprises carrying out the step of: h) performing a genome-wide association study (GWAS). In some

embodiments, the GWAS identifies single-nucleotide polymorphism (SNP) modifiers of the differential aging value in the plurality of individuals. In some embodiments, the GWAS identifies genetic modifiers of the phenotype in the plurality of individuals. In some embodiments, the haplotype of interest is defined as 0, 1, or 2 allele copies. In some embodiments, the allele copies are determined by SNP genotyping. In some embodiments, the phenotype correlated with a haplotype of interest is an expression level of a plurality of genes. In some embodiments, the gene expression level is the gene expression level in the brain. In some embodiments, the gene expression level is the gene expression level in the frontal cortex. In some embodiments, wherein the gene expression level is the gene expression level in the cerebellum.

[00180] In certain aspects, the invention provides an apparatus configured to determine the phenotype of a sample from a subject, wherein the phenotype is correlated with a haplotype of interest. In some embodiments, the apparatus is configured to a) provide a gene expression level of a plurality of genes in a sample for a plurality of individuals, wherein each individual's haplotype of interest is known; b) perform a linear regression of the expression level of each gene of the plurality of genes as a function of the haplotype of interest for each of the plurality of individuals to determine a linear regression co-efficient for each gene of the plurality of genes; c) select, from the plurality of genes, the genes whose expression level is significantly correlated with the haplotype of interest; d) determine, for each gene whose expression level is significantly correlated with the haplotype of interest selected in step (c), a residual value, wherein the residual value is determined by comparing the gene expression level of said gene in a sample from a subject to the linear regression performed in step (b) for said gene; e) determine, for each gene whose expression level is significantly correlated with the haplotype of interest, a ratio of the residual value determined in step (d) to the linear regression coefficient determined in step (b); f) repeat steps (d) and (e) for each gene whose expression level is significantly correlated with the haplotype of interest selected in step (c); and g) integrate the ratios for each gene whose expression level is significantly correlated with the haplotype of interest, wherein the integrated ratio corresponds to the phenotype of the sample from the subject.

[00181] In some embodiments, the plurality of individuals includes the subject. In some embodiments, steps (d) to (g) are performed for each individual in the plurality of individuals.

[00182] In some embodiments, the apparatus further comprises carrying out the step of: h) performing a genome-wide association study (GWAS). In some embodiments, the GWAS identifies single-nucleotide polymorphism (SNP) modifiers of the differential aging value in the plurality of individuals. In some embodiments, the GWAS identifies genetic modifiers of the phenotype in the plurality of individuals. In some embodiments, the haplotype of interest is defined as 0, 1, or 2 allele copies. In some embodiments, the allele copies are determined by SNP genotyping. In some embodiments, the phenotype correlated with a haplotype of interest is an expression level of a plurality of genes. In some embodiments, the gene expression level is the gene expression level in the brain. In some embodiments, the gene expression level is the gene expression level in the frontal cortex. In some embodiments, wherein the gene expression level is the gene expression level in the cerebellum.

[00183] In certain aspects, the invention provides a system comprising an apparatus configured to determine the phenotype of a sample from a subject, wherein the phenotype is correlated with a haplotype of interest. In some embodiments, the system comprises an apparatus configured to a) provide a gene expression level of a plurality of genes in a sample for a plurality of individuals, wherein each individual's haplotype of interest is known; b) perform a linear regression of the expression level of each gene of the plurality of genes as a function of the haplotype of interest for each of the plurality of individuals to determine a linear regression co-efficient for each gene of the plurality of genes; c) select, from the plurality of genes, the genes whose expression level is significantly correlated with the haplotype of interest; d) determine, for each gene whose expression level is significantly correlated with the haplotype of interest selected in step (c), a residual value, wherein the residual value is determined by comparing the gene expression level of said gene in a sample from a subject to the linear regression performed in step (b) for said gene; e) determine, for each gene whose expression level is significantly correlated with the haplotype of interest, a ratio of the residual value determined in step (d) to the linear regression coefficient determined in step (b); f) repeat steps (d) and (e) for each gene whose expression level is significantly correlated with the haplotype of interest selected in step (c); and g) integrate the ratios for each gene whose expression level is significantly correlated with the haplotype of interest, wherein the integrated ratio corresponds to the phenotype of the sample from the subject.

[00184] In some embodiments, the plurality of individuals includes the subject. In some embodiments, steps (d) to (g) are performed for each individual in the plurality of individuals.

[00185] In some embodiments, the system further comprises carrying out the step of: h) performing a genome-wide association study (GWAS). In some embodiments, the GWAS identifies single-nucleotide polymorphism (SNP) modifiers of the differential aging value in the plurality of individuals. In some embodiments, the GWAS identifies genetic modifiers of the phenotype in the plurality of individuals. In some embodiments, the haplotype of interest is defined as 0, 1, or 2 allele copies. In some embodiments, the allele copies are determined by SNP genotyping. In some embodiments, the phenotype correlated with a haplotype of interest is an expression level of a plurality of genes. In some embodiments, the gene expression level is the gene expression level in the brain. In some embodiments, the gene expression level is the gene expression level in the frontal cortex. In some embodiments, wherein the gene expression level is the gene expression level in the cerebellum.

[00186] In certain aspects, the invention provides a computer program product for identifying one or more genetic modifiers of a plurality of genes whose expression level is correlated with a genotype of interest comprising instructions which, when the program is executed by a computer, cause the computer to carry out the steps of: a) providing a gene expression level of a plurality of genes in a sample for a plurality of individuals, wherein each individual's genotype of interest is known; b) performing a linear regression of the expression level of each gene of the plurality of genes as a function of the genotype of interest of each of the plurality of individuals to determine a linear regression co-efficient for each gene of the plurality of genes; c) selecting, from the plurality of genes, the genes whose expression level is significantly correlated with the genotype of interest; d) determining, for each gene whose expression level is significantly correlated with the genotype of interest selected in step (c), a residual value, wherein the residual value is determined by comparing the gene expression level of said gene in a sample from a subject to the linear regression performed in step (b) for said gene; e) determining, for each gene whose expression level is significantly correlated with the genotype of interest, a ratio of the residual value determined in step (d) to the linear regression coefficient determined in step (b); f) repeating steps (d) and (e) for each gene whose expression level is significantly correlated with the genotype of interest selected in step (c); g) integrating the ratios for each gene significantly correlated with the genotype of interest; h) performing steps (d) to (g) for each individual in the plurality

of individuals; and j) identifying genetic modifiers that modulate the integrated ratio value determined in step (g).

[00187] In some embodiments, the genetic modifiers that modulate the integrated ratio value determined in step (g) are determined by performing a genome-wide association study (GWAS). In some embodiments, the plurality of individuals includes the subject. In some embodiments, the GWAS identifies single-nucleotide polymorphism (SNP) modifiers of the integrated ratio value determined in step (g) in the plurality of individuals. In some embodiments, the genetic modifier is a single-nucleotide polymorphism (SNP). In some embodiments, the genotype of interest is the risk allele of TMEM106B. In some embodiments, the risk allele of TMEM106B is and A at SNP rs1990622. In some embodiments, the genotype of interest is a risk allele associated with a disease or disorder. In some embodiments, the genotype of interest is a non-risk allele. In some embodiments, the genotype of interest is a haplotype.

[00188] In some embodiments, the gene expression level is the gene expression level in the brain. In some embodiments, the gene expression level is the gene expression level in the frontal cortex. In some embodiments, the gene expression level is the gene expression level in the cerebellum.

[00189] In some embodiments, the plurality of individuals are healthy individuals. In some embodiments, the plurality of individuals have a disease or disorder. In some embodiments, the plurality of individuals have a neurodegenerative disease. In some embodiments, the neurodegenerative disease is Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, or Fronto-temporal dementia.

[00190] In certain aspects, the invention provides an apparatus configured to identify one or more genetic modifiers of a plurality of genes whose expression level is correlated with a genotype of interest. In some embodiments, the apparatus is configured to a) provide a gene expression level of a plurality of genes in a sample for a plurality of individuals, wherein each individual's genotype of interest is known; b) perform a linear regression of the expression level of each gene of the plurality of genes as a function of the genotype of interest of each of the plurality of individuals to determine a linear regression co-efficient for each gene of the plurality of genes; c) select, from the plurality of genes, the genes whose expression level is significantly correlated with the genotype of interest; d) determine, for

each gene whose expression level is significantly correlated with the genotype of interest selected in step (c), a residual value, wherein the residual value is determined by comparing the gene expression level of said gene in a sample from a subject to the linear regression performed in step (b) for said gene; e) determine, for each gene whose expression level is significantly correlated with the genotype of interest, a ratio of the residual value determined in step (d) to the linear regression coefficient determined in step (b); f) repeat steps (d) and (e) for each gene whose expression level is significantly correlated with the genotype of interest selected in step (c); g) integrate the ratios for each gene significantly correlated with the genotype of interest; h) perform steps (d) to (g) for each individual in the plurality of individuals; and j) identify genetic modifiers that modulate the integrated ratio value determined in step (g).

[00191] In some embodiments, the genetic modifiers that modulate the integrated ratio value determined in step (g) are determined by performing a genome-wide association study (GWAS). In some embodiments, the plurality of individuals includes the subject. In some embodiments, the GWAS identifies single-nucleotide polymorphism (SNP) modifiers of the integrated ratio value determined in step (g) in the plurality of individuals. In some embodiments, the genetic modifier is a single-nucleotide polymorphism (SNP). In some embodiments, the genotype of interest is the risk allele of TMEM106B. In some embodiments, the risk allele of TMEM106B is and A at SNP rs1990622. In some embodiments, the genotype of interest is a risk allele associated with a disease or disorder. In some embodiments, the genotype of interest is a non-risk allele. In some embodiments, the genotype of interest is a haplotype.

[00192] In some embodiments, the gene expression level is the gene expression level in the brain. In some embodiments, the gene expression level is the gene expression level in the frontal cortex. In some embodiments, the gene expression level is the gene expression level in the cerebellum.

[00193] In some embodiments, the plurality of individuals are healthy individuals. In some embodiments, the plurality of individuals have a disease or disorder. In some embodiments, the plurality of individuals have a neurodegenerative disease. In some embodiments, the neurodegenerative disease is Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, or Fronto-temporal dementia.

[00194] In certain aspects, the invention provides a system comprising an apparatus configured to identify one or more genetic modifiers of a plurality of genes whose expression level is correlated with a genotype of interest. In some embodiments, the system comprises an apparatus configured to a) provide a gene expression level of a plurality of genes in a sample for a plurality of individuals, wherein each individual's genotype of interest is known; b) perform a linear regression of the expression level of each gene of the plurality of genes as a function of the genotype of interest of each of the plurality of individuals to determine a linear regression co-efficient for each gene of the plurality of genes; c) select, from the plurality of genes, the genes whose expression level is significantly correlated with the genotype of interest; d) determine, for each gene whose expression level is significantly correlated with the genotype of interest selected in step (c), a residual value, wherein the residual value is determined by comparing the gene expression level of said gene in a sample from a subject to the linear regression performed in step (b) for said gene; e) determine, for each gene whose expression level is significantly correlated with the genotype of interest, a ratio of the residual value determined in step (d) to the linear regression coefficient determined in step (b); f) repeat steps (d) and (e) for each gene whose expression level is significantly correlated with the genotype of interest selected in step (c); g) integrate the ratios for each gene significantly correlated with the genotype of interest; h) perform steps (d) to (g) for each individual in the plurality of individuals; and j) identify genetic modifiers that modulate the integrated ratio value determined in step (g).

[00195] In some embodiments, the genetic modifiers that modulate the integrated ratio value determined in step (g) are determined by performing a genome-wide association study (GWAS). In some embodiments, the plurality of individuals includes the subject. In some embodiments, the GWAS identifies single-nucleotide polymorphism (SNP) modifiers of the integrated ratio value determined in step (g) in the plurality of individuals. In some embodiments, the genetic modifier is a single-nucleotide polymorphism (SNP). In some embodiments, the genotype of interest is the risk allele of TMEM106B. In some embodiments, the risk allele of TMEM106B is and A at SNP rs1990622. In some embodiments, the genotype of interest is a risk allele associated with a disease or disorder. In some embodiments, the genotype of interest is a non-risk allele. In some embodiments, the genotype of interest is a haplotype.

[00196] In some embodiments, the gene expression level is the gene expression level in the brain. In some embodiments, the gene expression level is the gene expression level in the frontal cortex. In some embodiments, the gene expression level is the gene expression level in the cerebellum.

[00197] In some embodiments, the plurality of individuals are healthy individuals. In some embodiments, the plurality of individuals have a disease or disorder. In some embodiments, the plurality of individuals have a neurodegenerative disease. In some embodiments, the neurodegenerative disease is Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, or Fronto-temporal dementia.

[00198] *Gene expression levels*

[00199] In certain aspects, nucleic acids isolated from the subject's sample are sequenced to determine the gene expression and identify genotypes of interest. Several techniques known in the art can be used to detect or quantify DNA expression, RNA expression, or nucleic acid sequences, which include, but are not limited to, sequencing, hybridization, amplification, and/or binding to specific ligands. Methods to quantify nucleic acids from biological samples are known in the art. Any suitable method to quantify nucleic acids from biological samples are contemplated for use in the invention. In a non-limiting embodiment, gene expression can be measured using RT-PCR, qPCR, microarrays, or RNAseq. Other methods of measuring gene expression are known in the art.

[00200] Sequencing can be performed using techniques well known in the art, using automatic sequencers. Sequencing can be performed on the complete gene or on specific domains thereof.

[00201] For any of the gene transcripts to be quantified, suitable primers specific for a gene may be designed by known methods in the art. In other embodiments, the skilled artisan is able to modify the sequences of the above-described primers by addition and/or deletion of one or a few nucleotide(s) at the 3' and/or 5' end, for example but not limited to addition of nucleotides at the 5' end of a primer.

[00202] In certain embodiments, gene transcripts may be quantified using specific probes in the RT-qPCR. In certain embodiments, the probe is preferably labeled. Several probe systems have been described for specifically measuring amplification of a target

sequence. They are usually constituted of an oligonucleotide complementary to said target sequence, which is bonded to pairs of fluorophore groups or fluorophore/quenchers, such that hybridization of the probe to its target and the successive amplification cycles cause an increase or reduction in the total fluorescence of the mixture, depending on the case, proportional to the amplification of the target sequence.

[00203] Non-limiting examples of labeling systems that can be used to carry out kinetic PCR are the TaqMan™ (ABI.RTM.), the AmpliSensor™ (InGen), and the Sunrise™ (Oncor®, Appligene®) systems. The skilled artisan can choose amongst these systems or other any other labeling systems.

[00204] Apart from the primers and probe sequence, the skilled artisan can use general knowledge concerning quantitative RT-PCR in order to determine the other parameters for performing the method according to the invention, for example but not limited to, cycling parameters, and quantification having regard to a housekeeping gene. Examples of such parameters are well known in the art.

[00205] In other embodiments, gene transcripts can be quantified using nucleic acid microarrays and probes designed to detect specific transcripts. In other embodiments, gene transcripts can be quantified using RNA sequencing (RNA-seq) or whole transcriptome shotgun sequences (WTSS), which uses next generation sequencing to quantify RNA present in a biological sample. Methods of performing RNA-seq are known in the art.

[00206] Any suitable biological sample can be used to determine gene expression of the genotype of interest. In some embodiments, the biological sample can be taken from body fluid, such as urine, saliva, bone marrow, blood, and derivative blood products (sera, plasma, PBMC, circulating cells, circulating RNA). In some embodiments, the biological sample can be taken from a human subject, from an animal, or from a cell culture. The biological sample can be obtained in vivo, in vitro or ex vivo. Non-limiting examples of biological samples include blood, serum, plasma, cerebrospinal fluid, mucus, tissue, cells, and the like, or any combination thereof. In a non-limiting embodiment, the biological sample is blood. In a non-limiting embodiment, the biological sample is serum. In a non-limiting embodiment, the biological sample is plasma. In a non-limiting embodiment, the biological sample is cerebellum blood. In a non-limiting embodiment, the biological sample is a brain tissue sample. Any suitable method to isolate nucleic acids from biological samples

are contemplated for use in the invention. Biological samples for analysis are stored under suitable conditions. In non-limiting examples biological samples are kept at about 4°C. In non-limiting examples biological samples are kept at about -20°C. In non-limiting examples biological samples are kept at about -70-80°C.

[00207] *Methods of modifying a phenotype associated with a genetic risk allele, treating preventing, or delaying the onset of aging, and treating, preventing, or delaying the onset of cognitive decline*

[00208] In certain aspects, the invention provides a method of modifying a phenotype associated with a TMEM106B risk allele in a subject in need thereof, the method comprising administering an effective amount of an IL2 modulator, and IL2RA modulator, an IL2RB modulator, an IL2RG modulator, an IL15 modulator, an IL15RA modulator, or a combination thereof to the subject.

[00209] In certain aspects, the invention provides a method of treating, preventing, or delaying the onset of aging in a subject in need thereof, the method comprising administering an effective amount of an IL2 modulator, and IL2RA modulator, an IL2RB modulator, an IL2RG modulator, an IL15 modulator, an IL15RA modulator, a TMEM106B modulator, or a combination thereof to the subject.

[00210] In certain aspects, the invention provides a method of treating, preventing, or delaying the onset of cognitive decline in a subject in need thereof, the method comprising administering an effective amount of an IL2 modulator, and IL2RA modulator, an IL2RB modulator, an IL2RG modulator, an IL15 modulator, an IL15RA modulator, a TMEM106B modulator, or a combination thereof to the subject.

[00211] In some embodiments, the subject is administered an IL2 modulator. In some embodiments, the subject is administered an IL2RA modulator. In some embodiments, the subject is administered an IL2RB modulator. In some embodiments, the subject is administered an IL2RG modulator. In some embodiments, the subject is administered an IL15 modulator. In some embodiments, the subject is administered an IL15RA modulator.

[00212] In some embodiments, the subject is homozygous for the TMEM106B risk allele. In some embodiments, the subject is heterozygous for the TMEM106B risk allele. In some embodiments, the subject is homozygous for the TMEM106B protective allele. In some

embodiments, the TMEM106B risk allele is an A at SNP rs1990622. In some embodiments, the TMEM106B protective allele is a G at SNP rs1990622.

[00213] In some embodiments, the IL2RA modulator increases expression of a IL2RA protective allele, or decreases expression of a IL2RA risk allele, or a combination thereof. In some embodiments, the IL2RA protective allele is an A at SNP rs12722515. In some embodiments, the IL2RA risk allele is an C at SNP rs12722515.

[00214] In some embodiments, the modulation increases expression of a TMEM106B protective allele.

[00215] In some embodiments, the modulation decreases the expression of the TMEM106B risk allele. In some embodiments, the TMEM106B risk allele is an A at SNP rs1990622. In some embodiments, the TMEM106B protective allele is a G at SNP rs1990622.

[00216] In some embodiments, the phenotype associated with a TMEM106B risk allele is a plurality of genes, and their expression levels, associated with the TMEM106B risk allele. In some embodiments, the phenotype associated with a TMEM106B risk allele is reduced and a phenotype associated with a TMEM106B protective allele is increased.

[00217] *Modulators of the Invention*

[00218] As used herein, a modulator can be, but is not limited to, a compound that interacts with a gene, or protein, polypeptide, or peptide, and modulates its activity or its expression. Some non-limiting examples of modulators include peptides (such as peptide fragments comprising a polypeptide encoded by a gene, or antibodies or fragments thereof), small molecules, and nucleic acids (such as siRNA or antisense RNA specific for a nucleic acid). The modulator can either increase the activity or expression of a protein encoded by a gene, or the modulator can decrease the activity or expression of a protein encoded by a gene.

[00219] The modulator can be an antagonist (e.g., an inhibitor). Antagonists can be molecules which, decrease the amount or the duration of the activity of a protein. Antagonists and inhibitors include proteins, nucleic acids, antibodies, small molecules, or any other molecules which decrease the activity of a protein.

[00220] The modulator can be an agonist. Agonists of a protein can be molecules which, increase or prolong the activity of a protein, agonists include, but are not limited to, proteins, nucleic acids, small molecules, or any other molecules which activate a protein.

[00221] In one embodiment, a modulator can be a peptide fragment. Fragments include all possible amino acid lengths between and including about 8 and about 100 amino acids, for example, lengths between about 10 and about 100 amino acids, between about 15 and about 100 amino acids, between about 20 and about 100 amino acids, between about 35 and about 100 amino acids, between about 40 and about 100 amino acids, between about 50 and about 100 amino acids, between about 70 and about 100 amino acids, between about 75 and about 100 amino acids, or between about 80 and about 100 amino acids. These peptide fragments can be obtained commercially or synthesized via liquid phase or solid phase synthesis methods (Atherton *et al.*, (1989) *Solid Phase Peptide Synthesis: A Practical Approach*. IRL Press, Oxford, England). The peptide fragments can be isolated from a natural source, genetically engineered, or chemically prepared. These methods are well known in the art.

[00222] A modulator, for example, an agonist or antagonist, can be a protein such as an antibody (monoclonal, polyclonal, humanized, chimeric, or fully human), or a binding fragment thereof. An antibody fragment can be a form of an antibody other than the full-length form and includes portions or components that exist within full-length antibodies, in addition to antibody fragments that have been engineered. Antibody fragments can include, but are not limited to, single chain Fv (scFv), diabodies, Fv, and (Fab')₂, triabodies, Fc, Fab, CDR1, CDR2, CDR3, combinations of CDR's, variable regions, tetrabodies, bifunctional hybrid antibodies, framework regions, constant regions, and the like (*see*, Maynard *et al.*, (2000) *Ann. Rev. Biomed. Eng.* 2:339-76; Hudson (1998) *Curr. Opin. Biotechnol.* 9:395-402). Antibodies can be obtained commercially, custom generated, or synthesized against an antigen of interest according to methods established in the art (Janeway *et al.*, (2001) *Immunobiology*, 5th ed., Garland Publishing).

[00223] A modulator, for example, an agonist or antagonist, can be selected from the group comprising: siRNA; interfering RNA or RNAi; dsRNA; RNA Polymerase III transcribed DNAs; ribozymes; and antisense nucleic acids, which can be RNA, DNA, or an artificial nucleic acid. Antisense oligonucleotides, including antisense DNA, RNA, and DNA/RNA molecules, act to directly block the translation of mRNA by binding to targeted mRNA, and preventing protein translation. Antisense oligonucleotides of at least about 15 bases can be

synthesized, e.g., by conventional phosphodiester techniques (Dallas *et al.*, (2006) *Med. Sci. Monit.* 12(4):RA67-74; Kalota *et al.*, (2006) *Handb. Exp. Pharmacol.* 173:173-96; Lutzelburger *et al.*, (2006) *Handb. Exp. Pharmacol.* 173:243-59). Antisense nucleotide sequences include, but are not limited to: morpholinos, 2'-O-methyl polynucleotides, DNA, RNA and the like.

[00224] siRNA comprises a double stranded structure containing from about 15 to about 50 base pairs, for example from about 21 to about 25 base pairs, and having a nucleotide sequence identical or nearly identical to an expressed target gene or RNA within the cell. siRNA comprises a sense RNA strand and a complementary antisense RNA strand annealed together by standard Watson-Crick base-pairing interactions. The sense strand comprises a nucleic acid sequence which is substantially identical to a nucleic acid sequence contained within the target miRNA molecule. "Substantially identical" to a target sequence contained within the target mRNA refers to a nucleic acid sequence that differs from the target sequence by about 3% or less. The sense and antisense strands of the siRNA can comprise two complementary, single-stranded RNA molecules, or can comprise a single molecule in which two complementary portions are base-paired and are covalently linked by a single-stranded "hairpin" area. *See also*, McManus and Sharp (2002) *Nat Rev Genetics*, 3:737-47, and Sen and Blau (2006) *FASEB J.*, 20:1293-99, the entire disclosures of which are herein incorporated by reference.

[00225] The siRNA can be altered RNA that differs from naturally-occurring RNA by the addition, deletion, substitution and/or alteration of one or more nucleotides. Such alterations can include addition of non-nucleotide material, such as to the end(s) of the siRNA or to one or more internal nucleotides of the siRNA, or modifications that make the siRNA resistant to nuclease digestion, or the substitution of one or more nucleotides in the siRNA with deoxyribonucleotides. One or both strands of the siRNA can also comprise a 3' overhang. As used herein, a 3' overhang refers to at least one unpaired nucleotide extending from the 3'-end of a duplexed RNA strand. For example, the siRNA can comprise at least one 3' overhang of from 1 to about 6 nucleotides (which includes ribonucleotides or deoxyribonucleotides) in length, or from 1 to about 5 nucleotides in length, or from 1 to about 4 nucleotides in length, or from about 2 to about 4 nucleotides in length. For example, each strand of the siRNA can comprise 3' overhangs of dithymidylic acid ("TT") or diuridylic acid ("uu").

[00226] siRNA can be produced chemically or biologically, or can be expressed from a recombinant plasmid or viral vector (for example, see U.S. Patent No. 7,294,504 and U.S. Patent No. 7,422,896, the entire disclosures of which are herein incorporated by reference). Exemplary methods for producing and testing dsRNA or siRNA molecules are described in U.S. Patent Application Publication No. 2002/0173478 to Gewirtz, U.S. Patent Application Publication No. 2007/0072204 to Hannon *et al.*, and in U.S. Patent Application Publication No. 2004/0018176 to Reich *et al.*, the entire disclosures of which are herein incorporated by reference.

[00227] RNA polymerase III transcribed DNAs contain promoters, such as the U6 promoter. These DNAs can be transcribed to produce small hairpin RNAs in the cell that can function as siRNA or linear RNAs that can function as antisense RNA. A modulator, for example, an agonist or antagonist, can contain ribonucleotides, deoxyribonucleotides, synthetic nucleotides, or any suitable combination such that the target RNA and/or gene is inhibited. In addition, these forms of nucleic acid can be single, double, triple, or quadruple stranded. (see for example Bass (2001) *Nature*, 411, 428-429; Elbashir *et al.*, (2001) *Nature*, 411, 494-498; and PCT Publication Nos. WO 00/44895, WO 01/36646, WO 99/32619, WO 00/01846, WO 01/29058, WO 99/07409, WO 00/44914).

[00228] A modulator, for example, an agonist or antagonist, can be a small molecule that binds to a protein and disrupts its function, or conversely, enhances its function. Small molecules are a diverse group of synthetic and natural substances generally having low molecular weights. They can be isolated from natural sources (for example, plants, fungi, microbes and the like), are obtained commercially and/or available as libraries or collections, or synthesized. Candidate small molecules can be identified via *in silico* screening or high-throughput (HTP) screening of combinatorial libraries. Most conventional pharmaceuticals, such as aspirin, penicillin, and many chemotherapeutics, are small molecules, can be obtained commercially, can be chemically synthesized, or can be obtained from random or combinatorial libraries as described below (Werner *et al.*, (2006) *Brief Funct. Genomic Proteomic* 5(1):32-6).

[00229] Knowledge of the primary sequence of a molecule of interest, such as the amino acid sequence, and the similarity of that sequence with proteins of known function, can provide information as to the inhibitors or antagonists of the protein of interest in addition to agonists. Identification and screening of agonists and antagonists is further facilitated by

determining structural features of the protein, e.g., using X-ray crystallography, neutron diffraction, nuclear magnetic resonance spectrometry, and other techniques for structure determination. These techniques provide for the rational design or identification of agonists and antagonists.

[00230] *Pharmaceutical Compositions and Administration for Therapy*

[00231] Treatments of the invention can be administered to the subject once (e.g., as a single injection or deposition). Alternatively, treatments of the invention can be administered once or twice daily to a subject in need thereof for a period of from about two to about twenty-eight days, or from about seven to about ten days. Treatments of the invention can also be administered once or twice daily to a subject for a period of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 times per year, or a combination thereof. Furthermore, treatments of the invention can be co-administrated with another therapeutic. Where a dosage regimen comprises multiple administrations, the effective amount of the treatment(s) administered to the subject can comprise the total amount of the treatment(s) administered over the entire dosage regimen.

[00232] Treatments can be administered to a subject by any means suitable for delivering the treatment to cells of the subject, such as brain tissue or neuronal cells. For example, treatments can be administered by methods suitable to transfect cells. Transfection methods for eukaryotic cells are well known in the art, and include direct injection of a nucleic acid into the nucleus or pronucleus of a cell; electroporation; liposome transfer or transfer mediated by lipophilic materials; receptor mediated nucleic acid delivery, bioballistic or particle acceleration; calcium phosphate precipitation, and transfection mediated by viral vectors.

[00233] The compositions of this invention can be formulated and administered to reduce the symptoms by any means that produces contact of the active ingredient with the agent's site of action in the body of a subject, such as a human or animal (e.g., a dog, cat, or horse). They can be administered by any conventional means available for use in conjunction with pharmaceuticals, either as individual therapeutic active ingredients or in a combination of therapeutic active ingredients. They can be administered alone, but are generally administered with a pharmaceutical carrier selected on the basis of the chosen route of administration and standard pharmaceutical practice.

[00234] The treatments of the invention may be administered to a subject in an amount effective to treat or prevent. One of skill in the art can readily determine what will be an effective amount of the treatments of the invention to be administered to a subject, taking into account whether the modulator is being used prophylactically or therapeutically, and taking into account other factors such as the age, weight and sex of the subject, any other drugs that the subject may be taking, any allergies or contraindications that the subject may have, and the like. For example, an effective amount can be determined by the skilled artisan using known procedures, including analysis of titration curves established *in vitro* or *in vivo*. Also, one of skill in the art can determine the effective dose from performing pilot experiments in suitable animal model species and scaling the doses up or down depending on the subject's weight etc. Effective amounts can also be determined by performing clinical trials in individuals of the same species as the subject, for example starting at a low dose and gradually increasing the dose and monitoring the effects on a neurodegenerative disorder. Appropriate dosing regimens can also be determined by one of skill in the art without undue experimentation, in order to determine, for example, whether to administer the agent in one single dose or in multiple doses, and in the case of multiple doses, to determine an effective interval between doses.

[00235] A therapeutically effective dose of a treatment can depend upon a number of factors known to those of ordinary skill in the art. The dose(s) of the modulators can vary, for example, depending upon the identity, size, and condition of the subject or sample being treated, further depending upon the route by which the composition is to be administered, if applicable, and the effect which the practitioner desires the modulator to have upon the target of interest. These amounts can be readily determined by a skilled artisan. These amounts include, for example, mg or microgram (mg) amounts per kilogram (kg) of subject weight, such as about 0.25 mg/kg, about 0.5 mg/kg, about 1 mg/kg, about 2 mg/kg, about 3 mg/kg, about 4 mg/kg, about 5 mg/kg, about 6 mg/kg, about 7 mg/kg, about 8 mg/kg, about 9 mg/kg or about 10 mg/kg, or between about 0.25 mg/kg to 0.5 mg/kg, 0.5 mg/kg to 1 mg/kg, 1 mg/kg to 2 mg/kg, 2 mg/kg to 3 mg/kg, 3 mg/kg to 4 mg/kg, 4 mg/kg to 5 mg/kg, 5 mg/kg to 6 mg/kg, 6 mg/kg to 7 mg/kg, 7 mg/kg to 8 mg/kg, 8 mg/kg to 9 mg/kg, or 9 mg/kg to 10 mg/kg, or any range in between. These amounts also include a unit dose of a modulator, for example, mg or mg amounts, such as at least about 0.25 mg, 0.5 mg, 1 mg, 2 mg, 5 mg, 10 mg, 20 mg, 30 mg, 40 mg, 50 mg, 60 mg, 70 mg, 80 mg, 90 mg, 100 mg, 110 mg, 120 mg, 130 mg, 140 mg, 150 mg, 160 mg, 170 mg, 180 mg, 190 mg, 200 mg, 225 mg, 250 mg, 275

mg, 300 mg, 325 mg, 350 mg, 375 mg, 400 mg, 425 mg, 450 mg, 475 mg, 500 mg, 525 mg, 550 mg, 575 mg, 600 mg, 625 mg, 650 mg, 675 mg, 700 mg, 750 mg, 800 mg, 850 mg, 900 mg, or more. Any of the therapeutic applications described herein can be applied to any subject in need of such therapy, including, for example, a mammal such as a dog, a cat, a cow, a horse, a rabbit, a monkey, a pig, a sheep, a goat, or a human.

[00236] For modulators that are antagonists or inhibitors, or agonists, or activators, of a protein, or modulators that increase or decrease the expression of a gene or genes, the instructions would specify use of the pharmaceutical composition.

[00237] Pharmaceutical compositions for use in accordance with the invention can be formulated in conventional manner using one or more physiologically acceptable carriers or excipients. The therapeutic compositions of the invention can be formulated for a variety of routes of administration, including systemic and topical or localized administration. Techniques and formulations generally can be found in Remington's Pharmaceutical Sciences, Meade Publishing Co., Easton, Pa (20th Ed., 2000), the entire disclosure of which is herein incorporated by reference. For systemic administration, an injection is useful, including intramuscular, intravenous, intraperitoneal, and subcutaneous. For injection, the therapeutic compositions of the invention can be formulated in liquid solutions, for example in physiologically compatible buffers such as Hank's solution or Ringer's solution. In addition, the therapeutic compositions can be formulated in solid form and redissolved or suspended immediately prior to use. Lyophilized forms are also included. Pharmaceutical compositions of the present invention are characterized as being at least sterile and pyrogen-free. These pharmaceutical formulations include formulations for human and veterinary use.

[00238] According to the invention, a pharmaceutically acceptable carrier can comprise any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. The use of such media and agents for pharmaceutically active substances is well known in the art. Any conventional media or agent that is compatible with the active modulator can be used. Supplementary active modulators can also be incorporated into the compositions.

[00239] A pharmaceutical composition containing a modulator of the invention can be administered in conjunction with a pharmaceutically acceptable carrier, for any of the therapeutic effects discussed herein. Such pharmaceutical compositions can comprise, for

example antibodies directed to polypeptides encoded by genes of interest or variants thereof, or agonists and antagonists of a polypeptide encoded by a gene of interest. The compositions can be administered alone or in combination with at least one other agent, such as a stabilizing compound, which can be administered in any sterile, biocompatible pharmaceutical carrier including, but not limited to, saline, buffered saline, dextrose, and water. The compositions can be administered to a patient alone, or in combination with other agents, drugs or hormones.

[00240] A pharmaceutical composition of the invention is formulated to be compatible with its intended route of administration. Examples of routes of administration include parenteral, e.g., intravenous, intradermal, subcutaneous, oral (e.g., inhalation), transdermal (topical), transmucosal, and rectal administration. Solutions or suspensions used for parenteral, intradermal, or subcutaneous applications can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid; buffers such as acetates, citrates or phosphates and agents for the adjustment of tonicity such as sodium chloride or dextrose. pH can be adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide. The parenteral preparation can be enclosed in ampoules, disposable syringes or multiple dose vials made of glass or plastic.

[00241] Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersions. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor EM™ (BASF, Parsippany, N.J.) or phosphate buffered saline (PBS). In all cases, the composition must be sterile and should be fluid to the extent that easy syringability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, a pharmaceutically acceptable polyol like glycerol, propylene glycol, liquid polyethylene glycol, and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the

use of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, it can be useful to include isotonic agents, for example, sugars, polyalcohols such as mannitol, sorbitol, sodium chloride in the composition. Prolonged absorption of injectable compositions can be brought about by incorporating an agent which delays absorption, for example, aluminum monostearate and gelatin.

[00242] Sterile injectable solutions can be prepared by incorporating the modulator (e.g., a small molecule, peptide or antibody) in the required amount in an appropriate solvent with one or a combination of ingredients enumerated herein, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle which contains a basic dispersion medium and the required other ingredients from those enumerated herein. In the case of sterile powders for the preparation of sterile injectable solutions, examples of useful preparation methods are vacuum drying and freeze-drying which yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

[00243] Oral compositions generally include an inert diluent or an edible carrier. They can be enclosed in gelatin capsules or compressed into tablets. For the purpose of oral therapeutic administration, the active compound can be incorporated with excipients and used in the form of tablets, troches, or capsules. Oral compositions can also be prepared using a fluid carrier for use as a mouthwash, wherein the compound in the fluid carrier is applied orally and swished and expectorated or swallowed.

[00244] Pharmaceutically compatible binding agents, and/or adjuvant materials can be included as part of the composition. The tablets, pills, capsules, troches and the like can contain any of the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or lactose, a disintegrating agent such as alginic acid, Primogel, or corn starch; a lubricant such as magnesium stearate or sterotes; a glidant such as colloidal silicon dioxide; a sweetening agent such as sucrose or saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring.

[00245] Systemic administration can also be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be

permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration, detergents, bile salts, and fusidic acid derivatives. Transmucosal administration can be accomplished through the use of nasal sprays or suppositories. For transdermal administration, the active modulators are formulated into ointments, salves, gels, or creams as generally known in the art. In some embodiments, the modulator can be applied via transdermal delivery systems, which slowly releases the active modulator for percutaneous absorption. Permeation enhancers can be used to facilitate transdermal penetration of the active factors in the conditioned media. Transdermal patches are described in for example, U.S. Pat. No. 5,407,713; U.S. Pat. No. 5,352,456; U.S. Pat. No. 5,332,213; U.S. Pat. No. 5,336,168; U.S. Pat. No. 5,290,561; U.S. Pat. No. 5,254,346; U.S. Pat. No. 5,164,189; U.S. Pat. No. 5,163,899; U.S. Pat. No. 5,088,977; U.S. Pat. No. 5,087,240; U.S. Pat. No. 5,008,110; and U.S. Pat. No. 4,921,475.

[00246] Administration of the modulator is not restricted to a single route, but may encompass administration by multiple routes. For instance, exemplary administrations by multiple routes include, among others, a combination of intradermal and intramuscular administration, or intradermal and subcutaneous administration. Multiple administrations may be sequential or concurrent. Other modes of application by multiple routes will be apparent to the skilled artisan.

[00247] The modulators of the invention may be formulated into compositions for administration to subjects for the treatment and/or prevention. Such compositions may comprise the modulators of the invention in admixture with one or more pharmaceutically acceptable diluents and/or carriers and optionally one or more other pharmaceutically acceptable additives. The pharmaceutically-acceptable diluents and/or carriers and any other additives must be “acceptable” in the sense of being compatible with the other ingredients of the composition and not deleterious to the subject to whom the composition will be administered. One of skill in the art can readily formulate the modulators of the invention into compositions suitable for administration to subjects, such as human subjects, for example using the teaching a standard text such as Remington’s Pharmaceutical Sciences, 18th ed., (Mack Publishing Company: Easton, Pa., 1990), pp. 1635-36), and by taking into account the selected route of delivery.

[00248] Examples of diluents and/or carriers and/or other additives that may be used include, but are not limited to, water, glycols, oils, alcohols, aqueous solvents, organic

solvents, DMSO, saline solutions, physiological buffer solutions, peptide carriers, starches, sugars, preservatives, antioxidants, coloring agents, pH buffering agents, granulating agents, lubricants, binders, disintegrating agents, emulsifiers, binders, excipients, extenders, glidants, solubilizers, stabilizers, surface active agents, suspending agents, tonicity agents, viscosity-altering agents, carboxymethyl cellulose, crystalline cellulose, glycerin, gum arabic, lactose, magnesium stearate, methyl cellulose, powders, saline, sodium alginate. The combination of diluents and/or carriers and/or other additives used can be varied taking into account the nature of the active agents used (for example the solubility and stability of the active agents), the route of delivery (e.g. oral, parenteral, etc.), whether the agents are to be delivered over an extended period (such as from a controlled-release capsule), whether the agents are to be co-administered with other agents, and various other factors. One of skill in the art will readily be able to formulate the modulators for the desired use without undue experimentation.

[00249] The modulators of the invention may be administered to a subject by any suitable method that allows the agent to exert its effect on the subject *in vivo*. For example, the compositions may be administered to the subject by known procedures including, but not limited to, by oral administration, sublingual or buccal administration, parenteral administration, transdermal administration, via inhalation, via nasal delivery, vaginally, rectally, and intramuscularly. The modulators of the invention may be administered parenterally, or by epifascial, intracapsular, intracutaneous, subcutaneous, intradermal, intrathecal, intramuscular, intraperitoneal, intrasternal, intravascular, intravenous, parenchymatous, or sublingual delivery. Delivery may be by injection, infusion, catheter delivery, or some other means, such as by tablet or spray. In one embodiment, the modulators of the invention are administered to the subject by way of delivery directly to the brain tissue, such as by way of a catheter inserted into, or in the proximity of the subject's brain, or by using delivery vehicles capable of targeting the drug to the brain. For example, the modulators of the invention may be conjugated to or administered in conjunction with an agent that is targeted to the brain, or the spinal cord, such as an antibody or antibody fragment. In one embodiment, the modulators of the invention are administered to the subject by way of delivery directly to the tissue of interest, such as by way of a catheter inserted into, or in the proximity of the subject's tissue of interest, or by using delivery vehicles capable of targeting the drug to the brain, or the spinal cord, such as an antibody or antibody fragment.

[00250] For oral administration, a formulation of the modulators of the invention may be presented as capsules, tablets, powders, granules, or as a suspension or solution. The formulation may contain conventional additives, such as lactose, mannitol, cornstarch or potato starch, binders, crystalline cellulose, cellulose derivatives, acacia, cornstarch, gelatins, disintegrators, potato starch, sodium carboxymethylcellulose, dibasic calcium phosphate, anhydrous or sodium starch glycolate, lubricants, and/or or magnesium stearate.

[00251] For parenteral administration (*i.e.*, administration by through a route other than the alimentary canal), the modulators of the invention may be combined with a sterile aqueous solution that is isotonic with the blood of the subject. Such a formulation may be prepared by dissolving the active ingredient in water containing physiologically-compatible substances, such as sodium chloride, glycine and the like, and having a buffered pH compatible with physiological conditions, so as to produce an aqueous solution, then rendering the solution sterile. The formulation may be presented in unit or multi-dose containers, such as sealed ampoules or vials. The formulation may be delivered by injection, infusion, or other means known in the art.

[00252] For transdermal administration, the modulators of the invention may be combined with skin penetration enhancers, such as propylene glycol, polyethylene glycol, isopropanol, ethanol, oleic acid, *N*-methylpyrrolidone and the like, which increase the permeability of the skin to the modulators of the invention and permit the modulators to penetrate through the skin and into the bloodstream. The modulators of the invention also may be further combined with a polymeric substance, such as ethylcellulose, hydroxypropyl cellulose, ethylene/vinylacetate, polyvinyl pyrrolidone, and the like, to provide the composition in gel form, which are dissolved in a solvent, such as methylene chloride, evaporated to the desired viscosity and then applied to backing material to provide a patch.

[00253] In some embodiments, the modulators of the invention are provided in unit dose form such as a tablet, capsule or single-dose injection or infusion vial.

[00254] Various routes of administration and various sites of cell implantation can be utilized, such as, subcutaneous, intramuscular, or in brain tissue, or neuronal tissue, in order to introduce aggregated population of cells into a site of preference. Once implanted in a subject (such as a mouse, rat, or human), the aggregated cells can then treat or prevent a neurodegenerative disorder within the subject. In one embodiment, transfected cells (for

example, cells expressing a protein encoded by a gene) are implanted in a subject to treat or prevent Parkinson's Disease and/or lysosomal toxicity caused by LRRK2 kinase inhibitors within the subject. In other embodiments, the transfected cells are cells derived from brain tissue. In further embodiments, the transfected cells are neuronal cells. Aggregated cells (for example, cells grown in a hanging drop culture) or transfected cells (for example, cells produced as described herein) maintained for 1 or more passages can be introduced (or implanted) into a subject (such as a rat, mouse, dog, cat, human, and the like).

[00255] "Subcutaneous" administration can refer to administration just beneath the skin (i.e., beneath the dermis). Generally, the subcutaneous tissue is a layer of fat and connective tissue that houses larger blood vessels and nerves. The size of this layer varies throughout the body and from person to person. The interface between the subcutaneous and muscle layers can be encompassed by subcutaneous administration.

[00256] Administration of the cell aggregates is not restricted to a single route, but can encompass administration by multiple routes. For instance, exemplary administrations by multiple routes include, among others, a combination of intradermal and intramuscular administration, or intradermal and subcutaneous administration. Multiple administrations can be sequential or concurrent. Other modes of application by multiple routes will be apparent to the skilled artisan.

[00257] In other embodiments, this implantation method will be a one-time treatment for some subjects. In further embodiments of the invention, multiple cell therapy implantations will be required. In some embodiments, the cells used for implantation will generally be subject-specific genetically engineered cells. In another embodiment, cells obtained from a different species or another individual of the same species can be used. Thus, using such cells can require administering an immunosuppressant to prevent rejection of the implanted cells. Such methods have also been described in U.S. Patent Publication US 2004/0057937 and PCT Publication No. WO 2001/32840, and are hereby incorporated by reference.

[00258] *Gene Therapy and Protein Replacement Methods*

[00259] Delivery of nucleic acids into viable cells can be effected *ex vivo*, *in situ*, or *in vivo* by use of vectors, such as viral vectors (e.g., lentivirus, adenovirus, adeno-associated virus, or a retrovirus), or *ex vivo* by use of physical DNA transfer methods (e.g., liposomes or chemical treatments). Non-limiting techniques suitable for the transfer of nucleic acid into

mammalian cells *in vitro* include the use of liposomes, electroporation, microinjection, cell fusion, DEAE-dextran, and the calcium phosphate precipitation method (*See*, for example, Anderson, *Nature*, supplement to vol. 392, no. 6679, pp. 25-20 (1998)). Introduction of a nucleic acid or a gene encoding a polypeptide of the invention can also be accomplished with extrachromosomal substrates (transient expression) or artificial chromosomes (stable expression). Cells can also be cultured *ex vivo* in the presence of therapeutic compositions of the present invention in order to proliferate or to produce a desired effect on or activity in such cells. Treated cells can then be introduced *in vivo* for therapeutic purposes.

[00260] Nucleic acids can be inserted into vectors and used as gene therapy vectors. A number of viruses have been used as gene transfer vectors, including papovaviruses, *e.g.*, SV40 (Madzak *et al.*, 1992), adenovirus (Berkner, 1992; Berkner *et al.*, 1988; Gorziglia and Kapikian, 1992; Quantin *et al.*, 1992; Rosenfeld *et al.*, 1992; Wilkinson *et al.*, 1992; Stratford-Perricaudet *et al.*, 1990), vaccinia virus (Moss, 1992), adeno-associated virus (Muzyczka, 1992; Ohi *et al.*, 1990), herpesviruses including HSV and EBV (Margolskee, 1992; Johnson *et al.*, 1992; Fink *et al.*, 1992; Breakfield and Geller, 1987; Freese *et al.*, 1990), and retroviruses of avian (Biandypadhyay and Temin, 1984; Petropoulos *et al.*, 1992), murine (Miller, 1992; Miller *et al.*, 1985; Sorge *et al.*, 1984; Mann and Baltimore, 1985; Miller *et al.*, 1988), and human origin (Shimada *et al.*, 1991; Helseth *et al.*, 1990; Page *et al.*, 1990; Buchschacher and Panganiban, 1992). Non-limiting examples of *in vivo* gene transfer techniques include transfection with viral (*e.g.*, retroviral) vectors (*see* U.S. Patent No. 5,252,479, which is incorporated by reference in its entirety) and viral coat protein-liposome mediated transfection (Dzau *et al.*, *Trends in Biotechnology* 11:205-210 (1993), incorporated entirely by reference). For example, naked DNA vaccines are generally known in the art; *see* Brower, *Nature Biotechnology*, 16:1304-1305 (1998), which is incorporated by reference in its entirety. Gene therapy vectors can be delivered to a subject by, for example, intravenous injection, local administration (*see, e.g.*, U.S. Patent No. 5,328,470) or by stereotactic injection (*see, e.g.*, Chen, *et al.*, 1994, *Proc. Natl. Acad. Sci. USA* 91: 3054-3057). The pharmaceutical preparation of the gene therapy vector can include the gene therapy vector in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alternatively, where the complete gene delivery vector can be produced intact from recombinant cells, *e.g.*, retroviral vectors, the pharmaceutical preparation can include one or more cells that produce the gene delivery system.

[00261] For reviews of gene therapy protocols and methods see Anderson *et al.*, *Science* 256:808-813 (1992); U.S. Patent Nos. 5,252,479, 5,747,469, 6,017,524, 6,143,290, 6,410,010 6,511,847; and U.S. Patent Publication Nos. US2002/0077313 and US2002/00069, which are all hereby incorporated by reference in their entireties. For additional reviews of gene therapy technology, see Friedmann, *Science*, 244:1275-1281 (1989); Verma, *Scientific American*: 68-84 (1990); Miller, *Nature*, 357: 455-460 (1992); Kikuchi *et al.*, *J Dermatol Sci.* 2008 May;50(2):87-98; Isaka *et al.*, *Expert Opin Drug Deliv.* 2007 Sep;4(5):561-71; Jager *et al.*, *Curr Gene Ther.* 2007 Aug;7(4):272-83; Wachler *et al.*, *Nat Rev Genet.* 2007 Aug;8(8):573-87; Jensen *et al.*, *Ann Med.* 2007;39(2):108-15; Herweijer *et al.*, *Gene Ther.* 2007 Jan;14(2):99-107; Eliyahu *et al.*, *Molecules*, 2005 Jan 31;10(1):34-64; and Altaras *et al.*, *Adv Biochem Eng Biotechnol.* 2005; 99:193-260, all of which are hereby incorporated by reference in their entireties.

[00262] In some embodiments, the gene therapy is a CRISPR-based gene therapy. Introducing targeted modifications in the genome for therapeutic purposes, can require highly efficient systems that are able to alter the existing DNA pattern with great precision. The CRISPR/Cas9 type II system consists of the Cas9 nuclease and a single guide RNA (sgRNA or gRNA), which is a fusion of a CRISPR RNA (crRNA) and a trans-activating crRNA (tracrRNA) that binds Cas9 nuclease and directs it to a target sequence based on a complementary base-pairing rule. The target sequence must be adjacent to a protospacer-adjacent motif (PAM) consisting of a canonical NGG or NAG sequence. At the recognition site, a double-strand break (DSB) is generated that can be repaired by non-homologous end joining (NHEJ), resulting in small insertions or deletions usually associated with loss of function (knockdown/knockout). In the presence of an exogenous donor DNA, by a homologydirected recombination (HDR) mechanism, precise modifications can be achieved at the targeted site, resulting in gain of function (knockin). For additional review of CRISPR technology, see Chira *et al.*, *Mol. Thr. Nuc. Acids.* 2017, Vol. 7, p211, the contents of which is hereby incorporated by reference in its entirety.

[00263] Protein replacement therapy can increase the amount of protein by exogenously introducing wild-type or biologically functional protein by way of infusion. A replacement polypeptide can be synthesized according to known chemical techniques or can be produced and purified via known molecular biological techniques. Protein replacement therapy has been developed for various disorders. For example, a wild-type protein can be purified from

a recombinant cellular expression system (*e.g.*, mammalian cells or insect cells-*see* U.S. Patent No. 5,580,757 to Desnick *et al.*; U.S. Patent Nos. 6,395,884 and 6,458,574 to Selden *et al.*; U.S. Patent No. 6,461,609 to Calhoun *et al.*; U.S. Patent No. 6,210,666 to Miyamura *et al.*; U.S. Patent No. 6,083,725 to Selden *et al.*; U.S. Patent No. 6,451,600 to Rasmussen *et al.*; U.S. Patent No. 5,236,838 to Rasmussen *et al.* and U.S. Patent No. 5,879,680 to Ginns *et al.*), human placenta, or animal milk (*see* U.S. Patent No. 6,188,045 to Reuser *et al.*), or other sources known in the art. After the infusion, the exogenous protein can be taken up by tissues through non-specific or receptor-mediated mechanism.

[00264] A polypeptide encoded by a gene of interest can also be delivered in a controlled release system. For example, the polypeptide can be administered using intravenous infusion, an implantable osmotic pump, a transdermal patch, liposomes, or other modes of administration. In one embodiment, a pump can be used (*see, i.e.*, Langer, *supra*; Sefton, *CRC Crit. Ref. Biomed. Eng.* 14:201 (1987); Buchwald *et al.*, *Surgery* 88:507 (1980); Saudek *et al.*, *N. Engl. J. Med.* 321:574 (1989)). In another embodiment, polymeric materials can be used (*see* *Medical Applications of Controlled Release*, Langer and Wise (eds.), CRC Pres., Boca Raton, Fla. (1974); *Controlled Drug Bioavailability, Drug Product Design and Performance*, Smolen and Ball (eds.), Wiley, New York (1984); Ranger and Peppas, *J. Macromol. Sci. Rev. Macromol. Chem.* 23:61 (1983); *see also* Levy *et al.*, *Science* 228:190 (1985); During *et al.*, *Ann. Neurol.* 25:351 (1989); Howard *et al.*, *J. Neurosurg.* 71:105 (1989)). In yet another embodiment, a controlled release system can be placed in proximity of the therapeutic target thus requiring only a fraction of the systemic dose (*see, e.g.*, Goodson, in *Medical Applications of Controlled Release*, *supra*, vol. 2, pp. 115-138 (1984)). Other controlled release systems are discussed in the review by Langer (*Science* 249:1527-1533 (1990)).

[00265] Unless otherwise defined, 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. Exemplary methods and materials are described below, although methods and materials similar or equivalent to those described herein can also be used in the practice or testing of the present invention.

[00266] All publications and other references mentioned herein are incorporated by reference in their entirety, as if each individual publication or reference were specifically and individually indicated to be incorporated by reference. Publications and references cited herein are not admitted to be prior art.

EXAMPLES

[00267] Examples are provided herein to facilitate a more complete understanding of the invention. The following examples illustrate the exemplary modes of making and practicing the invention. However, the scope of the invention is not limited to specific embodiments disclosed in these Examples, which are for purposes of illustration only, since alternative methods can be utilized to obtain similar results.

[00268] Example 1 – Genetic determinants of aging in human brain

[00269] The progression of human age-associated traits, such as cognitive decline, is highly variable across the population, with some individuals appearing older or younger than expected at a given chronological age. Described herein is an unbiased method, termed Differential-aging (Δ -aging), that quantifies age-associated individual variability, and apply this approach to the analysis of transcriptome-wide cerebral cortex gene expression. A subsequent genome-wide association study in search of genetic modifiers of Δ -aging identified the TMEM106B gene locus, previously associated with frontotemporal dementia, as determinant of Δ -aging in cerebral cortex with genome-wide significance ($p < 10^{-20}$), in a meta-analysis of several cohorts totaling 1904 autopsied human brain samples. TMEM106B risk variants promote age-associated changes, such as inflammation, neuronal loss, and cognitive deficits, even in the absence of known brain disease. Surprisingly, the effect of the TMEM106B risk allele on Δ -aging is highly selective for the frontal cerebral cortex of older individuals (>65 yo). These data suggest a mechanistic link between accelerated brain aging and neurodegeneration.

[00270] Described herein is the identification of genetic and environmental factors that determine whether an individual appears younger or older than peers (Δ -aging). TMEM106B and GRN variants interact genetically in the regulation of Δ -aging. The role of TMEM106B in aging appears CNS region and life-stage selective. TMEM106B risk variants modulate CNS inflammatory and degenerative changes in the presence or absence of neurodegenerative disease.

[00271] Introduction

[00272] The rate at which human age-associated phenomena advance in otherwise healthy individuals, termed healthy biological aging, is highly variable (Deary et al., 2012; Jones et al., 2014; Pitt and Kaeberlein, 2015). This has been hypothesized to be a consequence, in part, of genetic heterogeneity across the population. However, specific genetic factors that determine the rate of normal biological aging remain to be elucidated. Rare Progeria syndromes are caused by single gene mutations, but these disorders are likely to be mechanistically distinct from the common healthy aging process (Burtner and Kennedy, 2010). Prior studies have also identified genetic factors such as apolipoprotein E that modify the likelihood of extreme longevity, as with centenarians, using genome-wide association studies (GWAS) (Deelen et al., 2014). But such extreme longevity may reflect a selective reduction in the incidence of some major causes of age-associated mortality, such as atherosclerosis, rather than an altered rate of biological aging.

[00273] The relationship between physiological “healthy” aging and aging-associated diseases is complex. Pathological hallmarks of Alzheimer’s disease (AD), which is a progressive dementia seen primarily in late life, include neurofibrillary tangles and amyloid plaques in the CNS, but these changes can also be found in the CNS of adults without clinical evidence of dementia, albeit to a lesser degree (Yu et al., 2015). Thus, it has been hypothesized that some aspect of healthy brain aging may represent a prodromal state to disease pathology. However, by functional criteria such as cognitive measures, changes associated with healthy aging appear distinct from those seen in neurodegeneration (Small et al., 2011).

[00274] Described herein is a quantitative approach, termed Differential-aging (Δ -aging), to evaluate the rate of aging across a cohort of tissue without prior assumption on the nature of the age-associated phenotypic changes, based on transcriptome-wide gene expression analyses. This Δ -aging trait reflects the difference between the apparent (“biological”) age of a tissue and the true (“chronological”) age of the individual from whom the sample was derived. Transcriptomic or epigenetic analyses have previously been used to identify age-associated phenotypic changes in a hypothesis-free manner (Bocklandt et al., 2011; Colantuoni et al., 2011; de Magalhaes et al., 2009; Glass et al., 2013; Hannum et al., 2013; Kang et al., 2011; Ori et al.; Zahn et al., 2007). However, most such changes are likely secondary to the aging process, rather than causal. To address causality, a genetic approach

is used. Specifically, Δ -aging analysis was applied to large-scale gene expression datasets from human brain tissue samples, and performed a genome-wide association study (GWAS). This approach identified common variants at 2 genetic loci, TMEM106B and Progranulin, associated with an increased rate of biological aging. The effect of TMEM106B risk variants was found to be selective to frontal cortex tissue in late life. Further annotation analyses revealed that the presence of TMEM106B risk variants leads to an increased inflammatory polarization of innate immune markers, and a reduction in neuronal markers. As the pro-inflammatory impact of the risk variants was seen even in the context of isolated innate immune cells, it was thought that this represents a proximal effect of the risk variant. Analysis of tissue from individuals with neurodegenerative diseases, including AD, suggest a broader role for TMEM106B in the CNS response to pathological or age-associated insults.

[00275] *Results*

[00276] *Data-driven quantification of biological aging*

[00277] In order to investigate potential determinants of the rate of biological aging, we defined a quantitative trait - termed Δ -aging - that captures whether an individual appears biologically younger or older than his or her true chronological age (Fig. 1A). The Δ -aging trait for a given individual within a cohort is a theoretical value defined as the difference between the apparent biological age and the true chronological age of the individual, and thus with the dimension of time (Fig. 1B).

[00278] To estimate the apparent biological age of a given individual within a cohort, we used transcriptome-wide gene expression data (but note that other biological datasets could similarly be used). Δ -aging analysis of transcriptomic data is performed in two steps (Figs. 8A-B, detailed in Supplementary Methods): (i) all transcripts that are correlated in their expression levels with the chronological ages of individuals within a given cohort of samples are identified; (ii) for each individual (or sample) within the cohort, the quantitative trait Δ -aging is defined as the difference between a predicted "biological age", that is based on the aggregate expression levels of the age-dependent transcripts, and the actual "chronological" age of the individual (Fig. 1B, Fig. 8). To address non-linear aspects of biologic aging in the most facile manner, cohorts may be subdivided into selected age ranges (such as late-life) to be studied independently.

[00279] In the most elementary case, Δ -aging for an individual may be quantified based on the analysis of a single gene whose expression level is significantly correlated with age within a given cohort (Fig. 1B, Fig. 8). However, such a limited analysis would most likely reflect gene-specific variations across the cohort, rather than an aspect of aging. Thus, to capture diverse age-associated phenotypes within a tissue of interest, Δ -aging herein represents an aggregated analysis of gene expression across the entire transcriptome of each individual (Fig. 8, see Supplementary Methods for details).

[00280] *Age-associated changes in gene expression patterns in human frontal cortex*

[00281] We initially sought to characterize the global effect of aging on transcriptome-wide gene expression in post-mortem, autopsied prefrontal cortex tissue samples from individuals free of known neuropathology (Fig. 2 A-C). The correlation between chronological age and gene expression levels was queried in transcriptome-wide microarray datasets from 4 independent cohorts (*Tgen* (Myers et al., 2007; Webster et al., 2009), *BrainEqtl* (Gibbs et al., 2010), *HBTRC* (Zhang et al., 2013) and *BrainCloud* (Colantuoni et al., 2011); detailed in Methods and Fig. 15). Meta-analysis of the results obtained in the 4 datasets (n=716 individuals >25 years old) identified 3329 genes that were significantly correlated in expression with chronological age (false discovery rate [FDR]<5% by linear regression, after correction for gender and batch effects, among the 10,474 genes present in all 4 datasets; see Methods for details; meta-analysis). Functional annotation revealed an age-associated decrease in the expression of neuronal genes and a parallel age-associated increase in the expression of genes characteristic of astrocytes, microglia and oligodendrocytes, as defined by the molecular signatures obtained by single-cell RNAseq from human brain (Darmanis et al., 2015) (Fig. 2A). These observations are consistent with the progressive age-associated loss of neurons and their processes, concomitant with astrogliosis and microglial expansion, as described in neuropathological studies (Beach et al., 1989; Mosher and Wyss-Coray, 2014).

[00282] A detailed examination of individual genes whose expression levels are positively or negatively correlated with chronological age -- such as the neuronal gene prepronociceptin (PNOC) or glial fibrillary acidic protein (GFAP) -- nonetheless revealed considerable variability between tissue samples obtained from different individuals at any given age; (Fig. 2B and 2C and meta-analysis). Consistent with the Δ -aging theoretical model presented above (Fig. 1B), we hypothesized that such variability may in part reflect biological diversity

in the aging process, due to genetic or non-genetic factors. In this model, for any given age-associated gene, some individual tissue samples may display expression levels that are higher or lower than expected for their chronological age, in part as a consequence of accelerated or decelerated biological aging (Fig. 1B).

[00283] *Genome-wide association study identifies TMEM106B as a genetic determinant of the rate of aging in human cerebral cortex*

[00284] Using Δ -aging as a quantitative trait, we next undertook a GWAS (Fig. 3A) in search of genetic modifiers of biological aging, in a meta-analysis across 4 transcriptome-wide frontal cortex gene expression datasets (N=716 individuals without known brain pathology, Fig. 15). A strong association was observed between Δ -aging and the single nucleotide polymorphism (SNP) variant rs1990622, at the TMEM106B gene locus in the Discovery cohort ($p=2.77 \times 10^{-7}$, $n=716$, Figs. 3B, Fig. 9), that was replicated with genome-wide significance in the Replication cohort ($p=1.68 \times 10^{-15}$, $n=497$, ROS-MAP cohort, see details below, Figs. 3B, Fig. 9), leading to a combined association p-value of 1.5×10^{-19} in 1213 samples, (Figs. 3B). As this association appeared most prominently within a dataset composed of older adults only (>65 yo; *Tgen* dataset; Fig. 3A, Fig. 15), we further refined the entire meta-analysis by stratifying individuals as either younger (<65 yo) or older (>65 yo) adults within all datasets. This approach revealed a striking age-dependence in the association between rs1990622 and Δ -aging, that reached genome-wide significance in the older but not the younger cohorts ($p=5.4 \times 10^{-10}$, $N=413$ and $p=8.0 \times 10^{-2}$, $N=303$ respectively; Fig. 3B, Figs. 8-9). Analysis of an independent, population-based RNA-Seq gene expression dataset of prefrontal cortex samples from older adults, that includes both unaffected individuals and individuals with neurological diseases, replicated the association between rs1990622 and Δ -aging (Religious Orders Study and Memory and Aging Project, *ROS-MAP* (Chan et al., 2015), $P=1.68 \times 10^{-15}$, $N=497$ samples from individuals >65 yo; Fig. 3B). Joint meta-analysis that included all of the above datasets yielded a combined p-value of 2.5×10^{-23} for the association between rs1990622 and Δ -aging ($n=910$ samples; Fig. 3C-E). In this meta-analysis, a second association was observed at SNP rs708384 on chromosome 17, which falls at the Progranulin gene locus, but this did not quite reach genome-wide statistical significance after correction for multiple testing (GRN; $p=6.23 \times 10^{-7}$ at rs708384, Fig. 3C). Taken together, these findings strongly support the hypothesis that genetic factors account for

some of the variability seen in the apparent rate of biological aging traits across the human population.

[00285] Strikingly, the TMEM106B SNP most strongly associated with increased Δ -aging in the meta-analysis, rs1990622, was previously also associated with FTD (Cruchaga et al., 2011; Finch et al., 2011; Van Deerlin et al., 2010). Specifically, the major allele at rs1990622 (A; “Risk”; with a 59.2% allelic frequency in population of European ascent), which in the meta-analysis above is associated to an increase in Δ -aging (Fig. 3A), had previously been shown to increase the risk of FTD (Van Deerlin et al., 2010), as well to reduce the age of onset comparably (Cruchaga et al., 2011; Finch et al., 2011). Similarly, local imputation at the GRN locus revealed that the strongest association with Δ -aging was observed for a variant (rs5848, $p=1.85 \times 10^{-8}$, Fig. 10) previously identified as associated with FTD as well as other neurodegenerative disorders (Chen et al., 2015) (Rademakers et al., 2008; van Blitterswijk et al., 2014) with the disease-associated risk allele being associated with an increase in Δ -aging. Thus, the same genetic determinants that modify the rate of apparent biological aging in the frontal cortex of otherwise healthy individuals also appear to play a significant role in a rare neurodegenerative disorder of the frontal cortex.

[00286] We next sought to determine whether the effect of TMEM106B genetic variation on the Δ -aging endophenotype (Fig. 3F, Fig. 11) might be reflected in functional measures of brain aging. We thus queried the relationship between TMEM106B haplotypes and cognitive function, as assessed by Mini Mental State Exam (MMSE) in a large cross-sectional cohort of genotyped individuals from the National Institute on Aging (NIA) Long Life Family Study (LLFS) (N=4953 individuals (Newman et al., 2011) (Barral et al., 2012)). Remarkably, in individuals without diagnosed dementia, the TMEM106B haplotype associated with a decreased rate of aging selectively in individuals over 65yo in the analysis above was also associated with better cognitive scores specifically in such older individuals ($p=2.8 \times 10^{-3}$, $n=774$, Fig. 4A-B; rs1060700 was used a proxy for rs1990622 as these are co-inherited, see Methods for details). Similar results were obtained in a second independent cross-sectional cohort (Health and Retirement Study; N=12507 genotyped individuals; (Juster and Suzman, 1995)) in which the TMEM106B haplotype associated with a decrease rate of aging in our analysis above was again associated with improved scores in a memory test specifically among individuals over 65yo but not younger individuals ($p=1.9 \times 10^{-2}$ in the >65yo age group, $n=5604$; $p=0.48$, in the younger individuals, $n=4432$, Fig. 4C). These analyses of

cognitive measures support the relevance of our transcriptomic-based Δ -aging studies of human frontal cortex tissue.

[00287] *The TMEM106B genetic variant modulates innate immune activation and neuronal loss markers*

[00288] To assess the impact of the TMEM106B rs1990622 risk allele load (defined as 0, 1 or 2 risk allele copies, as determined by SNP genotyping) on cerebral cortex gene expression in detail, we next compared the transcriptome-wide gene expression changes observed in the context of increasing risk allele load (termed the TMEM106B risk-associated transcriptomic signature of change) to the gene expression changes associated with chronological aging (termed the age-associated transcriptomic signature of change) in 4 independent gene expression datasets from prefrontal cortex tissue of individuals free of known neuropathology (*Tgen*, *BrainEqtl*, *HBTRC* and *BrainCloud*, Fig. 15). A meta-analysis demonstrated that the TMEM106B risk-associated transcriptomic signature of change was broadly correlated with the age-associated transcriptomic signature of change. Furthermore, this correlation appeared selective for tissue from older adults (>65 yo), relative to tissue from younger adults (Fig. 5A-C, meta-analysis).

[00289] To gain insight into molecular mechanisms that act downstream of the TMEM106B risk variant, we next applied a gene co-expression network approach, termed whole genome co-expression network analysis (WGCNA) (Fuller et al., 2007; Langfelder and Horvath, 2008), that can functionally probe transcriptomic patterns of change. WGCNA is a computational tool that clusters genes in an unsupervised (hypothesis-free) manner based on their correlated co-expression, and thus defines biologically relevant groups of genes that typically correspond to specific cell types or processes (Langfelder and Horvath, 2008). WGCNA analysis of a frontal cortical gene expression dataset from older adults defined 5 gene clusters (Fig. 12A; *Tgen* dataset; 179 individuals) that could be categorized in relation to major CNS cell types: microglia, astroglia, oligodendroglia, and 2 different neuron-associated groups. Assessment of the effect size of the rs1990622 genotype on the expression level of each of these gene clusters revealed the greatest impact to be on the microglia-associated gene cluster, which overall is significantly increased in expression with increased risk allele load. A similar pattern of transcriptome-wide gene expression change was observed with chronological aging, as expected. In contrast, other factors -- such as gender or post-mortem interval (PMI) to time of autopsy -- did not significantly affect these gene expression

categories in a consistent manner. In a complementary approach to WGCNA, we functionally annotated the TMEM106B risk-associated transcriptomic signature, as well as the age-associated transcriptomic signature, using previously described human CNS single-cell RNAseq data (Darmanis et al., 2015). This analysis confirmed the microglia gene set to be most increased in expression in the context of the rs1990622 risk variant; again, the effect was seen selectively in the older adult cohort (Fig. 12B, meta-analysis; N=716 unaffected individuals from *Tgen*, *BrainEqtl*, *HBTRC* and *BrainCloud* cohorts).

[00290] Prior studies have associated pathological aging with an inappropriate polarization of microglia and other innate immune myeloid cells towards an increasingly pro-inflammatory state (Deelen et al., 2014; Gabuzda and Yankner, 2013; Hu et al., 2015; Mosher and Wyss-Coray, 2014; Salminen et al., 2012). Given the altered expression of microglia-associated genes in the context of the TMEM106B risk haplotype, we hypothesized that this may be associated with a selective change in the expression of microglial inflammatory state-associated (“M1”) or anti-inflammatory/repair state-associated (“M2”) genes (Fig. 6A). Overall, both the M1- and M2-associated gene sets (detailed in Figs. 27A-B), as defined by RNAseq in polarized mouse microglia (Butovsky et al., 2014), were generally increased in expression over the course of chronological aging (Fig. 6B). TMEM106B rs1990622 risk-allele carriers showed a significantly muted age-associated increase in the expression of the M2 gene set (Fig. 6B, meta-analysis). M1 genes showed a trend towards a potentiated age-associated increase in expression that did not reach statistical significance in protective-allele carriers. Thus, taken together, the age-associated M1/M2 polarization changes appeared significantly shifted towards a pro-inflammatory state in cerebral cortex tissue from carriers of the TMEM106B risk allele.

[00291] We extended this analysis of inflammatory gene expression by additionally querying the effect of the TMEM106B rs1990622 genotype on another human myeloid cell lineage, dendritic cells. In a gene expression dataset from isolated, unstimulated human monocyte-derived dendritic cells from genotyped individuals without known pathology (GEO GSE53165 (Lee et al., 2014)), presence of the TMEM106B Δ -aging risk allele was again associated with an increased pro-inflammatory M1-like gene expression signature. Furthermore, the effect of the TMEM106B risk allele appeared non-additive with classical pro-inflammatory stimuli such as lipopolysaccharide (LPS) treatment (Fig. 6C, Fig. 13).

[00292] An unexpected observation in the analysis of this dendritic cell dataset (Lee et al., 2014) was that LPS exposure led to a significant reduction in the gene expression level of TMEM106B (Fig. 6D). A similar effect of LPS exposure on TMEM106B gene expression was seen in 2 additional gene expression datasets from other isolated human myeloid cell lineages, monocytes and macrophages (decrease >50%; GSE5099 and GSE10316, N=15 and 60, Fig. 5E-F). We further note that the TMEM106B rs1990622 risk allele was itself similarly associated with decreased expression of TMEM106B mRNA (Fig. 6D), suggesting a potential direct molecular mechanism for the effect of the TMEM106B risk haplotype on TMEM106B activity. Taken together, these findings suggest a regulatory circuitry, where TMEM106B activity impacts myeloid cell inflammatory status, and conversely where myeloid cell inflammatory status impacts TMEM106B activity. Those data suggest an “inflammaging” (Franceschi et al., 2007; Giunta et al., 2008) mechanism in which TMEM106B activity modulates the innate immune response in brain specifically in elderly individuals.

[00293] *TMEM106B, inflammaging and neurodegenerative disorders*

[00294] To further explore the relationships between Δ -aging and age-associated neurodegenerative diseases, we next analyzed frontal cortex gene expression datasets from individuals with a diagnosis of AD or Huntington’s disease (HD). Frontal cortex tissue from such individuals demonstrated significantly increased Δ -aging relative to unaffected individuals (Fig. 7A, plus 18.8 and 15.4 years respectively; HBTRC datasets), and thus appeared significantly older than expected in terms of their transcriptomic profiles. As this dataset also includes tissue samples from the cerebellum, we could further extend the analysis of Δ -aging to this second brain region. The effect of either AD or HD on Δ -aging appeared selective for frontal cortex, relative to cerebellar tissue, consistent with the neuropathological regional patterns that typify these disorders (Fig. 7A).

[00295] Further utilizing this cerebellar gene expression dataset, we extended the analysis of the association between the TMEM106B risk allele and Δ -aging to this brain region. The impact of the TMEM106B allele on Δ -aging was significantly less robust in cerebellar cortex tissue relative to frontal cortex, and a GWAS analysis in these datasets failed to identify an association of the TMEM106B alleles with Δ -aging in cerebellar tissue, in contrast to frontal cortex tissue (Fig. 14; in older individuals from the same cohorts).

[00296] We note that the effect of the TMEM106B risk allele appeared additive with the effect of AD on Δ -aging (Fig. 7A, Fig. 9), rather than occlusive, suggesting that the mechanisms are distinct. Consistent with this, previous studies have established a lack of association between TMEM106B and AD risk (Association p-values for rs1990622 of 0.26 and 0.87 in stage 1 and 2 respectively of the Alzheimer Disease Genetics Consortium (ADGC) GWAS in more than 22,000 individuals (Harold et al., 2009)). Furthermore, Apolipoprotein E (APOE) alleles -- which are major genetic determinants of AD risk -- were not associated with an alteration in Δ -aging (Fig. 28).

[00297] A broader analysis of 70 neurodegenerative disease-associated genetic variants (Jun et al., 2015; Lambert et al., 2013; McMillan et al., 2014; Nalls et al., 2014; Rollinson et al., 2011; van Es et al., 2009) did not reveal any significant additional associations with Δ -aging, beyond TMEM106B and GRN (Fig. 28). TMEM106B and GRN share a number of common attributes: both have previously been associated with risk of FTD (Cruchaga et al., 2011; Finch et al., 2011; Van Deerlin et al., 2010), with primary hippocampal sclerosis (Aoki et al., 2015), and with TDP-43 neuropathology in the absence of a clinical neurological diagnosis (Dickson et al., 2015; Yu et al., 2015). Furthermore, both genes have been implicated together in the regulation of lysosomal function (Schwenk et al., 2014; Stagi et al., 2014), and TMEM106B has been reported to regulate Progranulin protein accumulation (Chen-Plotkin et al., 2012). Consistent with a common pathway of action for these genes, the risk-associated genetic variants at these 2 loci showed a significant genetic interaction in their modulation of Δ -aging, in that the effect of GRN rs5848 variants on Δ -aging was observed only in carriers of the TMEM106B risk allele, where it reached genome-wide statistical significance ($p=1.91 \times 10^{-9}$ and 0.48 with $N=689$ and 187 in rs1990622 risk allele carriers and non-carrier respectively in Discovery+Replication cohorts, $p=6.42 \times 10^{-10}$ and 0.97 with $N=1137$ and 296 in Discovery+Replication_Disease cohorts, as defined in Fig. 3A and Fig. 15).

[00298] The fact that the TMEM106B genetic variant is not associated with AD risk and that AD genetic risk factors such as APOE4 do not appear associated with Δ -aging would suggest distinct phenomena. It moreover argues strongly against the possibility that we are merely observing a prodromal AD phenotype in the individuals with high Δ -aging values. The effect of the TMEM106B risk variant on Δ -aging in AD patients is significant, but as these patients display markers of accelerated aging, including inflammation and neuronal loss, even

independent of TMEM106B, the pathological relevance of TMEM106B is likely to be limited in this context.

[00299] *Non-genetic modifiers of aging*

[00300] To extend the application of our approach to aging beyond the study of genetic determinants, we further hypothesized that interventions previously associated with healthy aging, such as exercise, may decrease the apparent biological age, as quantified by Δ -aging, in contrast to the increased Δ -aging seen in pathological contexts such as AD (de Cabo et al., 2014). As a proof of principle, we analyzed existing datasets from a longitudinal gene expression study of serial human muscle tissue biopsies, sampled before and after vigorous exercise. Δ -aging analysis revealed that muscle tissue from individuals who undertook a vigorous exercise regimen for 6 months displayed a significant reduction in Δ -aging, and thus appeared significantly younger than their chronological age (Fig. 7B). Δ -aging may thus serve as a useful biomarker for the evaluation of anti-aging interventions.

[00301] *Discussion*

[00302] Δ -aging analysis allows for an unbiased quantification of an individual's apparent (biological) age, relative to other individuals within the same cohort. Δ -aging differs qualitatively from other aging analysis frameworks such as the "epigenetic clock" (Bocklandt et al., 2011; Lu et al., 2016), which assumes that aging impacts the expression of the same genes, and the same cellular processes, through all stages of life and in every tissue or context. By contrast, the Δ -aging analysis allows a context-dependent identification of age-associated genes, enabling tissue- or age- range-specific processes to be detected and taken into account. The absence of described association between TMEM106B or GRN and pleiotropic age-associated markers such as the "methylation clock" or telomere length suggests that TMEM106B and GRN impact the rate of healthy aging in prefrontal cortex independently of such factors. While we primarily used transcriptome-wide expression data to study aging in brain, subsequent studies may apply this approach to other tissues, and other data types and more systematically query the overlap with other aging-associated markers and the tissue-specificity of the effect of TMEM106B and GRN on healthy aging.

[00303] Our genome-wide association studies identified 2 genetic loci, at TMEM106B and Progranulin, that function cooperatively to modify Δ -aging in the cerebral cortex of older individuals. Our findings underscore the genetic differences between mechanisms that

govern healthy aging in a given tissue and those that underlie extreme longevity. Genetic variants previously associated with extreme longevity, such as APOE (Deelen et al., 2014; Deelen et al., 2011), were not associated with the rate of healthy aging or with age-associated inflammation in the present study.

[00304] The relationship between normal aging and neurodegenerative disorders appears complex. The 2 gene loci we identified as modifying Δ -aging have previously been associated with FTD risk. It appears extremely unlikely that our results are merely a reflection of “early onset FTD” in the cohorts we analyze for several reasons: First, FTD is an extremely rare disease. Its occurrence rate is 10 per 100,000 individuals in the 60-69 individuals (Knopman et al., 2004; Knopman and Roberts, 2011); the odds of the presence of contaminating case of FTD -- or even of prodromal FTD -- in any of the cohorts we studied are thus very low, while the association between TMEM106B genotype and Delta-aging is seen strongly in each and every of those cohorts. The fact that this association is observed exclusively in individuals aged of more than 65yo makes such an explanation even less likely, as FTD onset occurs typically between 40 and 65 yo (Knopman and Roberts, 2011). Furthermore, the “Unaffected individuals” of these cohorts were neuropathologically assessed; a FTD case would very likely have been identified and filtered out. A potential explanation for the dual association of those loci with both FTD risk and Delta-aging is that the proximate effect of these risk variants is to cause the tissue substrate for FTD -- frontal cortex -- to appear older, and thus secondarily to increase the prevalence of FTD (as prevalence of FTD is highly age-dependent). Although there are pathological criteria that differentiate healthy aging from disorders such as AD or FTD, certain processes are common, such as inflammation. Furthermore, neurodegenerative disease hallmarks, such as TDP-43 aggregates, are seen even in apparently healthy individuals, albeit to a limited extent (Beecham et al., 2014; Crary et al., 2014; Yu et al., 2015). Indeed, TMEM106B risk variants have been associated with increased TDP-43 aggregates in neuropathology-based association studies of apparently healthy older individuals (Dickson et al., 2015; Yu et al., 2015). We hypothesize that the selective role of TMEM106B in the aging frontal cortex may reflect unique stressors present in this tissue late in life, such as the accumulation of inflammatory cell debris or protein aggregates. The TMEM106B-Progranulin pathway may modulate the response to such stressors both during healthy aging and in the context of neurodegenerative disease (Fig. 6C) (Martens et al., 2012; Tanaka et al., 2013; Yin et al., 2010). Further studies in model systems may help to unravel underlying cellular mechanisms. Phenotypes

associated with TMEM106B genotype in myeloid cells (Fig. 6 A, C-D, Fig. 12) are of particular interest as they open the perspective of peripheral biomarkers of brain aging in monocytes or macrophages. Although our analysis focuses on the role of TMEM106B in inflammation, we do not exclude a function in neurons (Brady et al., 2013; Chen-Plotkin et al., 2012; Schwenk et al., 2014).

[00305] *Experimental Procedures*

[00306] *Gene expression analysis in human brain.* Gene expression datasets were downloaded from GEO (GEO GSE30272 and phs000417 for BrainEqtl and BrainCloud respectively), Synapse (syn4505 and syn3388564 for HBTRC and ROS-MAP respectively) or NIAGADS. All subsequent data manipulations and analyses were done using R statistical software. Within each dataset independently of the others, expression level matrixes were log-centered before further manipulation for homogenization. The effect of chronological age on each probe/gene was assessed using R *lm()* function, with gender, batch and post-mortem interval as correlates. The effect of rs1990622 genotype was similarly studied, with age, gender, batch and post-mortem interval as correlates. The effects are studied in either all individuals (age > 25 years old), in older adults (age > 65 years old for all datasets except BrainCloud for which individuals with age > 50 years old are included, as only n=9 individuals are aged of more than 65 years) and in younger adults (age > 25 years old and <65 years old for all datasets except BrainCloud for which individuals with age > 25 years old and <50 years old are included), as indicated in legends. Meta-analysis across different datasets were carried at a gene-level using *Stouffer's* weighted Z-score method.

[00307] *Δ -aging analysis.* For each individual sample an associated Δ -aging numerical value, expressed in time units is evaluated. It corresponds to the aggregation of Δ -aging values evaluated for each gene found to be significantly correlated with age in the studied dataset (with $\text{fdr}=5\%$). For such a given gene G, the gene-specific Δ -aging value in a sample from individual I corresponds to the difference between the age as it would be imputed on the sole basis of gene G expression level in the studied sample, and the actual chronological age of I. Formally, it is expressed as $\Delta_{GI} = \frac{\sigma_{GI}}{\sigma_G}$, where σ_G is the coefficient of the linear regression analysis of gene G expression levels in function of age across samples of the studied dataset and σ_{GI} the residual expression level of gene G in individual I in the

context of the above-mentioned linear regression. Integrating over the N genes with expression levels significantly correlated with age in the dataset of interest, the Δ -aging value

for individual I is expressed as $\Delta_I = \frac{1}{N} \sum_{G=1}^N \frac{\beta_{GI}}{\sigma_{GI}}$. Detailed explanation as well as a R script used for the calculation are provided in Supplemental Material and Methods.

[00308] *Genotype association analysis.* Genotypes datasets were downloaded from dbGap (phs000249 and GSE30272 for *BrainEqtl* and *BrainCloud* respectively), NIAGADS (NG00029 and NG0028 for *ROS-MAP* and *Tgen*) or Synapse (phs000417 for *HBTRC*). All subsequent data manipulations and analyses were done using PLINK 1.9 software (<https://www.cog-genomics.org/plink2>, (Chang et al., 2015)). Meta-analyses were carried using Metal software (Willer et al., 2010). Manhattan plots were drawn using R qqman package or LocusZoom (Pruim et al., 2010). Genotype imputation at targeted loci was performed using Impute2 software (Howie et al., 2009) and 1000 Genomes Phase I integrated haplotypes. For cognition association analysis, genotypes and cognitive assessment phenotypes datasets were downloaded from dbGap (phs000397 and phs000428 for Long Life Family Study and Health and Retirement Study respectively). Association between genotype and cognitive scores were tested using plink with the following covariates: age, gender and 3 population eigenvectors as defined by PCA.

[00309] *Gene co-expression analysis.* Unsupervised clustering was carried using R *WGCNA* package (Langfelder and Horvath, 2008) with the following settings: power = 8, TOMType = "signed", minModuleSize = 40, reassignThreshold = 0.05, mergeCutHeight = 0.25. Enrichment analysis for the identified clusters was done using WGCNA R package built-in *userListEnrichment* function and associated brain cell types categories.

[00310] *Human dendritic cell analysis:* Expression data from monocyte-derived dendritic cells from the Immvar cohort (Lee et al., 2014) were downloaded from GEO website (GSE57542); genotype for those cells were kindly shared by Dr. Philip L De Jager. Only individuals described as from European ascent were included in the analysis. The effects of treatment and rs1990622 on M1- or M2- related genes or on TMEM106B mRNA levels were studied by R by Kruskal-Wallis test or ANOVA, depending on the normality of the distribution and the similarity of variance between the groups as assessed using Shapiro and Levene tests respectively.

[00311] *Transcriptome-Wide Aging Meta-Analysis*. Transcriptome-wide meta-analysis results for the effect of age on gene expression levels in neurodegenerative-disease free human prefrontal cortex samples, carried in 4 datasets: *Tgen*, *BrainEqtl*, *HBTRC* and *BrainCloud*, described in Fig. 3B was performed. Analysis for all individuals of age >25yo, for older adults only with age >65yo or for younger adults only 25yo<Age<65yo was performed.

[00312] *Transcriptome-Wide rs1990622 risk allele load Meta-Analysis*. Transcriptome-wide meta-analysis results for the effect of TMEM106B rs1990622 risk allele (A) load on gene expression levels in neurodegenerative-disease free human prefrontal cortex samples, carried in 4 datasets: *Tgen*, *BrainEqtl*, *HBTRC* and *BrainCloud*, described in Fig. 3B was performed. Analysis for all individuals of age >25yo, for older adults only with age >65yo or for younger adults only 25yo<Age<65yo was performed.

[00313] *Transcriptome-Wide Aging Meta-Analysis Stratified by TMEM106B rs1990622*. Transcriptome-wide meta-analysis results for the effect of age on gene expression levels in neurodegenerative-disease free human prefrontal cortex samples, carried in 4 datasets: *Tgen*, *BrainEqtl*, *HBTRC* and *BrainCloud*, described in Fig. 3B and stratified for TMEM106B rs1990622 genotype was performed. Analysis for individuals of age >25yo in the 3 genotype groups: risk allele homozygotes (RR), protective allele homozygotes (PP) or heterozygotes (PR) for the rs1990622 variant was performed.

[00314] *References*

Aoki, N., Murray, M.E., Ogaki, K., Fujioka, S., Rutherford, N.J., Rademakers, R., Ross, O.A., and Dickson, D.W. (2015). Hippocampal sclerosis in Lewy body disease is a TDP-43 proteinopathy similar to FTLD-TDP Type A. *Acta Neuropathol* 129, 53-64.

Barral, S., Cosentino, S., Costa, R., Matteini, A., Christensen, K., Andersen, S.L., Glynn, N.W., Newman, A.B., and Mayeux, R. (2012). Cognitive function in families with exceptional survival. *Neurobiol Aging* 33, 619 e611-617.

Beach, T.G., Walker, R., and McGeer, E.G. (1989). Patterns of gliosis in alzheimer's disease and aging cerebrum. *Glia* 2, 420-436.

- Beecham, G.W., Hamilton, K., Naj, A.C., Martin, E.R., Huentelman, M., Myers, A.J., Comeveaux, J.J., Hardy, J., Vonsattel, J.P., Younkin, S.G., *et al.* (2014). Genome-wide association meta-analysis of neuropathologic features of Alzheimer's disease and related dementias. *PLoS Genet* 10, e1004606.
- Bocklandt, S., Lin, W., Sehl, M.E., Sanchez, F.J., Sinsheimer, J.S., Horvath, S., and Vilain, E. (2011). Epigenetic predictor of age. *PLoS One* 6, e14821.
- Brady, O.A., Zheng, Y., Murphy, K., Huang, M., and Hu, F. (2013). The frontotemporal lobar degeneration risk factor, TMEM106B, regulates lysosomal morphology and function. *Hum Mol Genet* 22, 685-695.
- Burtner, C.R., and Kennedy, B.K. (2010). Progeria syndromes and ageing: what is the connection? *Nat Rev Mol Cell Biol* 11, 567-578.
- Butovsky, O., Jedrychowski, M.P., Moore, C.S., Cialic, R., Lanser, A.J., Gabriely, G., Koeglisperger, T., Dake, B., Wu, P.M., Doykan, C.E., *et al.* (2014). Identification of a unique TGF-beta-dependent molecular and functional signature in microglia. *Nat Neurosci* 17, 131-143.
- Chan, G., White, C.C., Winn, P.A., Cimpean, M., Replogle, J.M., Glick, L.R., Cuerdon, N.E., Ryan, K.J., Johnson, K.A., Schneider, J.A., *et al.* (2015). CD33 modulates TREM2: convergence of Alzheimer loci. *Nat Neurosci* 18, 1556-1558.
- Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., and Lee, J.J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 4, 7.
- Chen-Plotkin, A.S., Unger, T.L., Gallagher, M.D., Bill, E., Kwong, L.K., Volpicelli-Daley, L., Busch, J.I., Akle, S., Grossman, M., Van Deerlin, V., *et al.* (2012). TMEM106B, the risk gene for frontotemporal dementia, is regulated by the microRNA-132/212 cluster and affects progranulin pathways. *J Neurosci* 32, 11213-11227.
- Chen, Y., Li, S., Su, L., Sheng, J., Lv, W., Chen, G., and Xu, Z. (2015). Association of progranulin polymorphism rs5848 with neurodegenerative diseases: a meta-analysis. *J Neurol* 262, 814-822.

Colantuoni, C., Lipska, B.K., Ye, T., Hyde, T.M., Tao, R., Leek, J.T., Colantuoni, E.A., Elkahoul, A.G., Herman, M.M., Weinberger, D.R., *et al.* (2011). Temporal dynamics and genetic control of transcription in the human prefrontal cortex. *Nature* 478, 519-523.

Crary, J.F., Trojanowski, J.Q., Schneider, J.A., Abisambra, J.F., Abner, E.L., Alafuzoff, I., Arnold, S.E., Attems, J., Beach, T.G., Bigio, E.H., *et al.* (2014). Primary age-related tauopathy (PART): a common pathology associated with human aging. *Acta Neuropathol* 128, 755-766.

Cruchaga, C., Graff, C., Chiang, H.H., Wang, J., Hinrichs, A.L., Spiegel, N., Bertelsen, S., Mayo, K., Norton, J.B., Morris, J.C., *et al.* (2011). Association of TMEM106B gene polymorphism with age at onset in granulin mutation carriers and plasma granulin protein levels. *Arch Neurol* 68, 581-586.

Darmanis, S., Sloan, S.A., Zhang, Y., Enge, M., Caneda, C., Shuer, L.M., Hayden Gephart, M.G., Barres, B.A., and Quake, S.R. (2015). A survey of human brain transcriptome diversity at the single cell level. *Proc Natl Acad Sci U S A* 112, 7285-7290.

de Cabo, R., Carmona-Gutierrez, D., Bernier, M., Hall, M.N., and Madeo, F. (2014). The search for antiaging interventions: from elixirs to fasting regimens. *Cell* 157, 1515-1526.

de Magalhaes, J.P., Curado, J., and Church, G.M. (2009). Meta-analysis of age-related gene expression profiles identifies common signatures of aging. *Bioinformatics* 25, 875-881.

Deary, I.J., Yang, J., Davies, G., Harris, S.E., Tenesa, A., Liewald, D., Luciano, M., Lopez, L.M., Gow, A.J., Corley, J., *et al.* (2012). Genetic contributions to stability and change in intelligence from childhood to old age. *Nature* 482, 212-215.

Deelen, J., Beekman, M., Uh, H.W., Broer, L., Ayers, K.L., Tan, Q., Kamatani, Y., Bennet, A.M., Tamm, R., Trompet, S., *et al.* (2014). Genome-wide association meta-analysis of human longevity identifies a novel locus conferring survival beyond 90 years of age. *Hum Mol Genet* 23, 4420-4432.

Deelen, J., Beekman, M., Uh, H.W., Helmer, Q., Kuningas, M., Christiansen, L., Kremer, D., van der Breggen, R., Suchiman, H.E., Lakenberg, N., *et al.* (2011). Genome-wide association study identifies a single major locus contributing to survival into old age; the APOE locus revisited. *Aging Cell* 10, 686-698.

Dickson, D.W., Rademakers, R., Nicholson, A.M., Schneider, J.A., Yu, L., and Bennett, D.A. (2015). The TMEM106B locus and TDP-43 pathology in older persons without FTLD. *Neurology* 85, 1354-1355.

Finch, N., Carrasquillo, M.M., Baker, M., Rutherford, N.J., Coppola, G., Dejesus-Hernandez, M., Crook, R., Hunter, T., Ghidoni, R., Benussi, L., *et al.* (2011). TMEM106B regulates progranulin levels and the penetrance of FTLD in GRN mutation carriers. *Neurology* 76, 467-474.

Franceschi, C., Capri, M., Monti, D., Giunta, S., Olivieri, F., Sevini, F., Panourgia, M.P., Invidia, L., Celani, L., Scurti, M., *et al.* (2007). Inflammaging and anti-inflammaging: a systemic perspective on aging and longevity emerged from studies in humans. *Mech Ageing Dev* 128, 92-105.

Fuller, T.F., Ghazalpour, A., Aten, J.E., Drake, T.A., Lusis, A.J., and Horvath, S. (2007). Weighted gene coexpression network analysis strategies applied to mouse weight. *Mamm Genome* 18, 463-472.

Gabuzda, D., and Yankner, B.A. (2013). Physiology: Inflammation links ageing to the brain. *Nature* 497, 197-198.

Gibbs, J.R., van der Brug, M.P., Hernandez, D.G., Traynor, B.J., Nalls, M.A., Lai, S.L., Arepalli, S., Dillman, A., Rafferty, I.P., Troncoso, J., *et al.* (2010). Abundant quantitative trait loci exist for DNA methylation and gene expression in human brain. *PLoS Genet* 6, e1000952.

Giunta, B., Fernandez, F., Nikolic, W.V., Obregon, D., Rrapo, E., Town, T., and Tan, J. (2008). Inflammaging as a prodrome to Alzheimer's disease. *J Neuroinflammation* 5, 51.

Glass, D., Vinuela, A., Davies, M.N., Ramasamy, A., Parts, L., Knowles, D., Brown, A.A., Hedman, A.K., Small, K.S., Buil, A., *et al.* (2013). Gene expression changes with age in skin, adipose tissue, blood and brain. *Genome Biol* 14, R75.

Hannum, G., Guinney, J., Zhao, L., Zhang, L., Hughes, G., Sada, S., Klotzle, B., Bibikova, M., Fan, J.B., Gao, Y., *et al.* (2013). Genome-wide methylation profiles reveal quantitative views of human aging rates. *Mol Cell* 49, 359-367.

Harold, D., Abraham, R., Hollingworth, P., Sims, R., Gerrish, A., Hamshere, M.L., Pabwa, J.S., Moskvina, V., Dowzell, K., Williams, A., *et al.* (2009). Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. *Nat Genet* 41, 1088-1093.

Howie, B.N., Donnelly, P., and Marchini, J. (2009). A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* 5, e1000529.

Hu, X., Leak, R.K., Shi, Y., Suenaga, J., Gao, Y., Zheng, P., and Chen, J. (2015). Microglial and macrophage polarization-new prospects for brain repair. *Nat Rev Neurol* 11, 56-64.

Jones, O.R., Scheuerlein, A., Salguero-Gomez, R., Camarda, C.G., Schaible, R., Casper, B.B., Dahlgren, J.P., Ehrlén, J., Garcia, M.B., Menges, E.S., *et al.* (2014). Diversity of ageing across the tree of life. *Nature* 505, 169-173.

Jun, G., Ibrahim-Verbaas, C.A., Vronskaya, M., Lambert, J.C., Chung, J., Naj, A.C., Kunkle, B.W., Wang, L.S., Bis, J.C., Bellenguez, C., *et al.* (2015). A novel Alzheimer disease locus located near the gene encoding tau protein. *Mol Psychiatry*.

Juster, F.T., and Suzman, R. (1995). An Overview of the Health and Retirement Study. *The Journal of Human Resources* 30, S7.

Kang, H.J., Kawasawa, Y.I., Cheng, F., Zhu, Y., Xu, X., Li, M., Sousa, A.M., Pletikos, M., Meyer, K.A., Sedmak, G., *et al.* (2011). Spatio-temporal transcriptome of the human brain. *Nature* 478, 483-489.

Knopman, D.S., Petersen, R.C., Edland, S.D., Cha, R.H., and Rocca, W.A. (2004). The incidence of frontotemporal lobar degeneration in Rochester, Minnesota, 1990 through 1994. *Neurology* 62, 506-508.

Knopman, D.S., and Roberts, R.O. (2011). Estimating the number of persons with frontotemporal lobar degeneration in the US population. *J Mol Neurosci* 45, 330-335.

Lambert, J.C., Ibrahim-Verbaas, C.A., Harold, D., Naj, A.C., Sims, R., Bellenguez, C., DeStafano, A.L., Bis, J.C., Beecham, G.W., Grenier-Boley, B., *et al.* (2013). Meta-analysis

of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet* 45, 1452-1458.

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 9, 559.

Lee, M.N., Ye, C., Villani, A.C., Raj, T., Li, W., Eisenhaure, T.M., Imboywa, S.H., Chipendo, P.I., Ran, F.A., Slowikowski, K., *et al.* (2014). Common genetic variants modulate pathogen-sensing responses in human dendritic cells. *Science* 343, 1246980.

Lu, A.T., Hannon, E., Levine, M.E., Hao, K., Crimmins, E.M., Lunnon, K., Kozlenkov, A., Mill, J., Dracheva, S., and Horvath, S. (2016). Genetic variants near MLST8 and DHX57 affect the epigenetic age of the cerebellum. *Nat Commun* 7, 10561.

Martens, L.H., Zhang, J., Barmada, S.J., Zhou, P., Kamiya, S., Sun, B., Min, S.W., Gan, L., Finkbeiner, S., Huang, E.J., *et al.* (2012). Progranulin deficiency promotes neuroinflammation and neuron loss following toxin-induced injury. *J Clin Invest* 122, 3955-3959.

McMillan, C.T., Toledo, J.B., Avants, B.B., Cook, P.A., Wood, E.M., Suh, E., Irwin, D.J., Powers, J., Olm, C., Elman, L., *et al.* (2014). Genetic and neuroanatomic associations in sporadic frontotemporal lobar degeneration. *Neurobiol Aging* 35, 1473-1482.

Mosher, K.I., and Wyss-Coray, T. (2014). Microglial dysfunction in brain aging and Alzheimer's disease. *Biochem Pharmacol* 88, 594-604.

Myers, A.J., Gibbs, J.R., Webster, J.A., Rohrer, K., Zhao, A., Marlowe, L., Kaleem, M., Leung, D., Bryden, L., Nath, P., *et al.* (2007). A survey of genetic human cortical gene expression. *Nat Genet* 39, 1494-1499.

Nalls, M.A., Pankratz, N., Lill, C.M., Do, C.B., Hernandez, D.G., Saad, M., DeStefano, A.L., Kara, E., Bras, J., Sharma, M., *et al.* (2014). Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. *Nat Genet* 46, 989-993.

Newman, A.B., Glynn, N.W., Taylor, C.A., Sebastiani, P., Perls, T.T., Mayeux, R., Christensen, K., Zmuda, J.M., Barral, S., Lee, J.H., *et al.* (2011). Health and function of

participants in the Long Life Family Study: A comparison with other cohorts. *Aging (Albany NY)* 3, 63-76.

Ori, A., Toyama, Brandon H., Harris, Michael S., Bock, T., Iskar, M., Bork, P., Ingolia, Nicholas T., Hetzer, Martin W., and Beck, M. Integrated Transcriptome and Proteome Analyses Reveal Organ-Specific Proteome Deterioration in Old Rats. *Cell Systems* 1, 224-237.

Pitt, J.N., and Kaeblerlein, M. (2015). Why is aging conserved and what can we do about it? *PLoS Biol* 13, e1002131.

Pruim, R.J., Welch, R.P., Sanna, S., Teslovich, T.M., Chines, P.S., Gliedt, T.P., Boehnke, M., Abecasis, G.R., and Willer, C.J. (2010). LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* 26, 2336-2337.

Rademakers, R., Eriksen, J.L., Baker, M., Robinson, T., Ahmed, Z., Lincoln, S.J., Finch, N., Rutherford, N.J., Crook, R.J., Josephs, K.A., *et al.* (2008). Common variation in the miR-659 binding-site of GRN is a major risk factor for TDP43-positive frontotemporal dementia. *Hum Mol Genet* 17, 3631-3642.

Rollinson, S., Mead, S., Snowden, J., Richardson, A., Rohrer, J., Halliwell, N., Usher, S., Neary, D., Mann, D., Hardy, J., *et al.* (2011). Frontotemporal lobar degeneration genome wide association study replication confirms a risk locus shared with amyotrophic lateral sclerosis. *Neurobiol Aging* 32, 758 e751-757.

Salminen, A., Kaarniranta, K., and Kauppinen, A. (2012). Inflammaging: disturbed interplay between autophagy and inflammasomes. *Aging (Albany NY)* 4, 166-175.

Schwenk, B.M., Lang, C.M., Hogg, S., Tabirovic, S., Orozco, D., Rentzsch, K., Lichtenthaler, S.F., Hoogenraad, C.C., Capell, A., Haass, C., *et al.* (2014). The FTL risk factor TMEM106B and MAP6 control dendritic trafficking of lysosomes. *EMBO J* 33, 450-467.

Small, S.A., Schobel, S.A., Buxton, R.B., Witter, M.P., and Barnes, C.A. (2011). A pathophysiological framework of hippocampal dysfunction in ageing and disease. *Nat Rev Neurosci* 12, 585-601.

- Stagi, M., Klein, Z.A., Gould, T.J., Bewersdorf, J., and Strittmatter, S.M. (2014). Lysosome size, motility and stress response regulated by fronto-temporal dementia modifier TMEM106B. *Mol Cell Neurosci* 61, 226-240.
- Tanaka, Y., Matsuwaki, T., Yamanouchi, K., and Nishihara, M. (2013). Exacerbated inflammatory responses related to activated microglia after traumatic brain injury in progranulin-deficient mice. *Neuroscience* 231, 49-60.
- van Blitterswijk, M., Mullen, B., Wojtas, A., Heckman, M.G., Diehl, N.N., Baker, M.C., DeJesus-Hernandez, M., Brown, P.H., Murray, M.E., Hsiung, G.Y., *et al.* (2014). Genetic modifiers in carriers of repeat expansions in the C9ORF72 gene. *Mol Neurodegener* 9, 38.
- Van Deerlin, V.M., Sleiman, P.M., Martinez-Lage, M., Chen-Plotkin, A., Wang, L.S., Graff-Radford, N.R., Dickson, D.W., Rademakers, R., Boeve, B.F., Grossman, M., *et al.* (2010). Common variants at 7p21 are associated with frontotemporal lobar degeneration with TDP-43 inclusions. *Nat Genet* 42, 234-239.
- van Es, M.A., Veldink, J.H., Saris, C.G., Blauw, H.M., van Vught, P.W., Birve, A., Lemmens, R., Schelhaas, H.J., Groen, E.J., Huisman, M.H., *et al.* (2009). Genome-wide association study identifies 19p13.3 (UNC13A) and 9p21.2 as susceptibility loci for sporadic amyotrophic lateral sclerosis. *Nat Genet* 41, 1083-1087.
- Webster, J.A., Gibbs, J.R., Clarke, J., Ray, M., Zhang, W., Holmans, P., Rohrer, K., Zhao, A., Marlowe, L., Kaleem, M., *et al.* (2009). Genetic control of human brain transcript expression in Alzheimer disease. *Am J Hum Genet* 84, 445-458.
- Willer, C.J., Li, Y., and Abecasis, G.R. (2010). METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 26, 2190-2191.
- Yin, F., Banerjee, R., Thomas, B., Zhou, P., Qian, L., Jia, T., Ma, X., Ma, Y., Iadecola, C., Beal, M.F., *et al.* (2010). Exaggerated inflammation, impaired host defense, and neuropathology in progranulin-deficient mice. *J Exp Med* 207, 117-128.
- Yu, L., De Jager, P.L., Yang, J., Trojanowski, J.Q., Bennett, D.A., and Schneider, J.A. (2015). The TMEM106B locus and TDP-43 pathology in older persons without FTL. *Neurology* 84, 927-934.

Zahn, J.M., Poosala, S., Owen, A.B., Ingram, D.K., Lustig, A., Carter, A., Weeraratna, A.T., Taub, D.D., Gorospe, M., Mazan-Mamczarz, K., *et al.* (2007). AGEMAP: a gene expression database for aging in mice. *PLoS Genet* 3, e201.

Zhang, B., Gaiteri, C., Bodea, L.G., Wang, Z., McElwee, J., Podtelezhnikov, A.A., Zhang, C., Xie, T., Tran, L., Dobrin, R., *et al.* (2013). Integrated systems approach identifies genetic nodes and networks in late-onset Alzheimer's disease. *Cell* 153, 707-720.

[00315] *Supplemental Experimental Procedures*

[00316] Theory and calculation of Delta-Age.

[00317] Principle

[00318] At the level of a gene whose expression level is positively correlated with age within a given cohort, we define the Delta Age of a given individual as the difference between the actual (chronological) age and the biological age as is would be imputed for this individual on the basis of gene G expression level data across the entire cohort.

[00319] Fig. 16 presents a theoretical case for illustration. Individuals, represented as dots, are plotted as a function of their chronological age (X-axis) and their measured expression level for gene G (Y-axis). The dotted line corresponds to the regression line of Gene G expression levels as a function of chronological age across the entire cohort. A graphical interpretation shows that dots that are above this regression line (in red) correspond to individuals with expression levels of G higher than expected for their chronological age, while those below (in blue) to individual with G levels lower than expected for their chronological age.

[00320] In Fig. 16, the biological age imputed on the basis of gene G expression level is presented for 2 individuals by green arrows, corresponding to the projection on the age axis through the regression line. The yellow lines correspond to the chronological age, and the Delta-age represents the difference.

[00321] Formal calculation of Delta-Age for a given individual and a single gene.

[00322] The linear regression across individuals of the expression level of a gene G ($Expr_G$) in function of chronological age ($ChrAge$) yields a regression line defined by the following equation:

$$(1) Expr_G = a_G \cdot ChrAge + b_G$$

[00323] Where a_G and b_G are respectively the factor and constant associated to gene G expression levels in the linear regression in function of age.

[00324] As graphically illustrated in Fig. 17, for a given individual I the expression level of gene G, $Expr_{GI}$ can be expressed by both expression:

(2a) $Expr_{GI} = a_G \cdot ChrAge_{GI} + b_G + \epsilon_{GI}$, where ϵ_{GI} is the residual expression level of gene G in individual I after the above-mentioned linear regression

And

(2b) $Expr_{GI} = a_G \cdot AppAge_{GI} + b_G$ where $AppAge_{GI}$ is the apparent age of individual I for gene G.

[00325] As a consequence:

$$(3) \Delta_{GI} = AppAge_{GI} - ChrAge_{GI} = \frac{\epsilon_{GI}}{a_G}$$

The Delta-Age for an individual I for a given gene G is expressed as the ratio between the residual value for the individual and the coefficient obtained by linear regression of the expression level of gene G in function of Age across individuals.

[00326] Formal calculation of Delta-Age across multiple genes (Fig. 18)

[00327] For a given individual I, the global Delta-age Δ_I is obtained by integration of all the gene-specific Delta-Age Δ_{GI} over all genes which expression levels are found to be correlated with chronological age during the original linear regression:

$$(4) \Delta_I = \frac{1}{N} \sum_{G=1}^N \Delta_{GI} = \frac{1}{N} \sum_{G=1}^N \frac{\epsilon_{GI}}{a_{GI}}$$

[00328] *Computation of Differential-Aging using R, annotated code.*

[00329] Delta calculation

[00330] ##Delta Calculation using a gene expression matrix *ExprNumLog* with expression probes as rows, samples as columns and expression levels in a log-scale.

[00331] ## Age and Gender (or additional covariates) are provided as vectors corresponding to the organization of the samples in the expression matrix columns.

[00332] ## *Coord_Temp* defines which samples –defined by their column coordinates - are to be included in the analysis. By default, all samples are included:

[00333] *Coord_Temp*<-1:length(*ExprNumLog*[1,])

[00334] ##The expression matrix is normed and centered. Its values correspond to $\frac{E_{ij} - \bar{E}_i - \bar{E}_j + \bar{E}}{\sqrt{(\sigma_i^2 + \sigma_j^2)}}$ in equations (2a) and (2b) above.

[00335] *Mtemp*<-CentreNormeDoubleGenesSdInd(*ExprNumLog*[,*Coord_Temp*])

[00336] ##A temporary matrix *Genes_ResidualsTemp* is created to store the residual values for each probe/sample pair after linear regression analysis for age and gender (and other potential experimental cofactors such as *Pmi*, *Batch*, ...). This matrix has the same size and structure as *ExprNumLog*. Its values correspond to the $\hat{\epsilon}_{ij}$ in equations (2a), (3) and (4) above.

Genes_ResidualsTemp<-matrix(0,length(*Mtemp*[,1]),length(*Coord_Temp*))

[00337] ##3 matrices are created to store for each probe the effect of age and gender. *Factor_AgeGender*, *Stat_AgeGender* and *Pval_AgeGender* will respectively store in columns 1/2 the estimated coefficient, t-statistic and corresponding p-value of the association with age/gender for the expression level of each probe, as determined by R's *lm()* function summary. Those matrices have the same number of rows as *ExprNumLog* (1 per probe) and 2 columns (Age/Gender effects). The values stored in *Factor_AgeGender* correspond to the β_{ij} in equations (1), (2a), (2b), (3) and (4) above.

Pval_AgeGender<-matrix(NA,length(*Mtemp*[,1]),2)

Factor_AgeGender<-matrix(NA,length(*Mtemp*[,1]),2)

Stat_AgeGender<-matrix(NA,length(*Mtemp*[,1]),2)

[00338] ##Each probe/row of the centered-normed expression matrix *Mtemp* is queried for the linear association of its levels with Age and Gender using R's *lm()* function. The results

are stored in *Genes_ResidualsTemp*, *Pval_AgeGender*, *Stat_AgeGender* and *Factor_AgeGender*.

```
for (i in 1:length(Mtemp[,1]))
```

```
{
```

[00339] ##The linear regression below corresponds to the one described in equation (1) above, with the addition of Gender as a covariate. Other covariates could be added at this stage (Batch, Pmi...).

```
y<-lm(Mtemp[i,]~Age[Coord_Temp]+Gender[Coord_Temp],na.action=na.exclude)
```

```
Genes_ResidualsTemp[i,]<-residuals(y)
```

```
Pval_AgeGender[i,1]<-t(summary(y)[[4]][2:3,4])
```

```
Stat_AgeGender[i,1]<-t(summary(y)[[4]][2:3,3])
```

```
Factor_AgeGender[i,1]<-t(summary(y)[[4]][2:3,1])
```

```
}
```

[00340] ##For each sample, the Delta-Age value is calculated using the residual expression levels after linear regression for age and gender. This corresponds to equation (4) above. The genes included in the calculation are the one found to be significantly associated with age in the regression analysis. 3 false-discovery rate cut-off thresholds for the inclusion of genes are considered: $fdr=1\%$, 5% or 10% . Delta-Age values are calculated independently for each of them.

[00341] ## The expression *SeuilPos(1-p.adjust(Pval_AgeGender[,1], method = "fdr"),0.99)* is used to select the genes that are associated with age with $fdr=1\%$: it returns a vector which length equals the number of probes in the original expression matrix, with binary values of 1 if the probe's level is linearly correlated with age or 0 if not, for the considered significance level.

```
DeltaAge_Div_Age_Factors_FDR1pc<-
```

```
as.numeric(t(t(Genes_ResidualsTemp)%*%(SeuilPos(1-p.adjust(Pval_AgeGender[,1],
```

```
method = "fdr"),0.99)/Factor_Age[,1])/sum(SeuilPos(1-p.adjust(Pval_AgeGender[,1],
```

```
method = "fdr"),0.99))))
```

```
DeltaAge_Div_Age_Factors_FDR5pc<-
```

```
as.numeric(t(t(Genes_ResidualsTemp)%*%(SeuilPos(1-p.adjust(Pval_AgeGender[,1],
```

```
method = "fdr"),0.95)/Factor_AgeGender[,1])/sum(SeuilPos(1-p.adjust(Pval_AgeGender[,1],
```

```

method = "fdr"),0.95))))
DeltaAge_Div_Age_Factors_FDR10pc<-
as.numeric(t(t(Genes_ResidualsTemp)%*%(SeuilPos(1-p.adjust(Pval_AgeGender[,1],
method = "fdr"),0.9)/Factor_AgeGender[,1])/sum(SeuilPos(1-p.adjust(Pval_AgeGender[,1],
method = "fdr"),0.9))))

```

[00342] ## Object such as *DeltaAge_Div_Age_Factors_FDR5pc* are the final output, being vector of length equal to the number of samples included in the analysis, containing the Differential-Aging values for each individual. Such values are later used as quantitative trait in genetic analysis.

[00343] ##Delta Calculation from a gene expression matrix *ExprNumLin* with expression probes as rows, samples as columns and expression levels in a linear-scale. The process is similar to the one described into detail above, except for the creation of *Mtemp* at line #2, for which the function *CentreNormeDoubleGenesLogSdInd* is used instead of *CentreNormeDoubleGenesSdInd*, with the consequence to first log-transforming of the expression matrix before norming/centering it.

[00344] ##NOTE: due to the log-transformation step, linear expression level matrix can contain only strictly positive values. Among the strategies to enforce this: 1) Rows/probes containing zero/negative values can be filtered out, 2) The whole expression matrix can be offset by its minimal value +1 (*ExprNumLin*<- *ExprNumLin*+*min(ExprNumLin)*+1), ...

```

Coord_Temp<-1:length(ExprNumLin[,1])
Mtemp<-CentreNormeDoubleGenesLogSdInd(ExprNumLin[,Coord_Temp])
Genes_ResidualsTemp<-matrix(0,length(Mtemp[,1]),length(Coord_Temp))
Pval_AgeGender<-matrix(NA,length(Mtemp[,1]),2)
Factor_AgeGender<-matrix(NA,length(Mtemp[,1]),2)
Stat_AgeGender<-matrix(NA,length(Mtemp[,1]),2)
for (i in 1:length(Mtemp[,1]))
{
y<-lm(Mtemp[i,]~Age[Coord_Temp],na.action=na.exclude)
Genes_ResidualsTemp[i,]<-residuals(y)
Pval_AgeGender[i,1]<-t(summary(y))[[4]][2,4])
Stat_AgeGender[i,1]<-t(summary(y))[[4]][2,3])
}

```

```

Factor_AgeGender[i,1]<-t(summary(y)[[4]][2,1])
}
DeltaAge_Div_Age_Factors_FDR1pc<-
as.numeric(t(t(Genes_ResidualsTemp)%*%(SeuilPos(1-p.adjust(Pval_AgeGender[,1],
method = "fdr"),0.99)/Factor_AgeGender[,1])/sum(SeuilPos(1-p.adjust(Pval_AgeGender[,1],
method = "fdr"),0.99))))
DeltaAge_Div_Age_Factors_FDR5pc<-
as.numeric(t(t(Genes_ResidualsTemp)%*%(SeuilPos(1-p.adjust(Pval_AgeGender[,1],
method = "fdr"),0.95)/Factor_AgeGender[,1])/sum(SeuilPos(1-p.adjust(Pval_AgeGender[,1],
method = "fdr"),0.95))))
DeltaAge_Div_Age_Factors_FDR10pc<-
as.numeric(t(t(Genes_ResidualsTemp)%*%(SeuilPos(1-p.adjust(Pval_AgeGender[,1],
method = "fdr"),0.9)/Factor_AgeGender[,1])/sum(SeuilPos(1-p.adjust(Pval_AgeGender[,1],
method = "fdr"),0.9))))

```

[00345] *Custom dependent functions:*

```

[00346] ## CentreNormeDoubleGenesSdInd
CentreNormeDoubleGenesSdInd<-function (Matrix)
{
  Mout <- Matrix
  for (i in 1:length(Matrix[, 1])) {

    A <- sum(Mout[i, ])/length(Mout[i, ])
    Mout[i, ] <- (Mout[i, ] - A)
  }
  for (j in 1:length(Matrix[1, ])) {

    B <- sum(Mout[,j])/length(Mout[,j])
    Mout[,j] <- (Mout[,j] - B)/sd(Mout[,j])
  }

  Mout
}

```

```

[00347] ## CentreNormeDoubleGenesLogSdInd
CentreNormeDoubleGenesLogSdInd<-function (Matrix)
{
  Mout <- Matrix
  for (i in 1:length(Matrix[, 1])) {
    Mout[i, ] <- log((Matrix[i, ]), 2)
    A <- sum(Mout[i, ])/length(Mout[i, ])
    Mout[i, ] <- (Mout[i, ] - A)
  }
  for (j in 1:length(Matrix[1, ])) {

    B <- sum(Mout[,j])/length(Mout[,j])
    Mout[,j] <- (Mout[,j] - B)/sd(Mout[,j])
  }

  Mout
}

```

```

[00348] ##SeuilPos

```

```

SeuilPos<-function(x,s)
{m<-max(x,na.rm=TRUE)
Pos(ceiling((x-s)/m))}

```

```

[00349] ##Pos

```

```

Pos<-function(x)
{(x+abs(x))/2}

```

```

[00350] Example 2

```

[00351] The delta-aging method can be applied to other tissues in a more systematic manner, for example, but not limited to, cerebellum blood. Genetic determinants of the sub-components of Delta-Aging can be determined. Detailed analysis of the effect of TMEM106B and HHIP can be performed such as determining the transcriptomic effect of the SNPs, experimental OE/KD, and genetic interactors for Delta modulation. Extension towards

other phenotypes can be performed (for example ADNI imaging, microRNA, methylation, proteins) either to expand the scope of application or to combine with RNA (noise reduction).

[00352] Fig. 19 shows the aging rate as a differential in an age-related trait. In red: individuals with a level higher than one would expect for their age: “looking older.” In blue: individuals with a level lower than one would expect for their age: “looking younger.”

[00353] Fig. 20 shows aging as a differential expression trait. In red: individuals with an expression level higher than one would expect for their age: “apparently older.” In blue: individuals with an expression level lower than one would expect for their age: “apparently older.”

[00354] Fig. 21 shows evaluating a delta age for a given gene.

[00355] Fig. 22 shows the model - principle of aging as a complex expression trait. Left graph: Gene positively associated with age (expression level increasing with age). Center graph: Gene not associated with age. Right graph: Gene negatively associated with age (expression level decreasing with age).

[00356] Fig. 23 shows aging as a complex expression trait. Combination across all the genes associated with age for a given individual is achieved by integrating all the genes affected by aging.

[00357] Fig. 24 shows the delta-age in 2 gene expression datasets in a tissue affected by Alzheimer’s Disease (prefrontal cortex). AD samples –here used as proxies for accelerated aged samples - display higher Deltas.

[00358] Fig. 25 shows the effect of diet in mice on delta-age (left). Effect of exercise in human muscle on delta-age (right).

[00359] Fig. 26 shows genetic determinants of aging rate in brain. Transcriptome-wide expression data in brain cortex samples from genotyped, neurodegenerative-diseases free individuals.

[00360] Stages to compute Δ_i in a given expression dataset:

[00361] 1) For each of the genes, proceed to a linear regression of its expression level (log-scale) in function of age and other cofactors (gender, pmi, batch...)

[00362] Provides: Residuals $\sigma_{G,I}$ for each individual/gene pair, and linear regression coefficient a_G for each gene.

[00363] 2) Select the genes whose level is significantly associated with aging, to be included in the Delta calculation.

[00364] Provides: N genes to be included in Δ_I calculation.

[00365] 3) Integrate for each individual the $\frac{\sigma_{G,I}}{a_G}$ terms across the selected genes.

[00366] Provides: Δ_I for each individual.

[00367] Aging as a complex expression trait is determined using the following expression:

$$\Delta_I = \frac{1}{N} \sum_{G=1}^N \frac{\sigma_{G,I}}{a_G}$$

[00368]

[00369] where:

[00370] Δ_I : Delta Age for individual I.

[00371] $\sigma_{G,I}$: Residual for individual I of a linear fit of G levels in function of age across individuals.

[00372] a_G : Linear regression coefficient of a linear fit of G levels in function of age across individuals.

[00373] As described, the Delta approach combines for each sample the effect of several genes whose levels are affected by age into a uni-dimensional factor that reflects an excess (in one direction or another) in the age-related transcriptional signature by comparison to the one expected for the sample's age, interpreted as over- or under- aging.

[00374] The delta is expressed in the same time unit as the input age (years, months, weeks...).

[00375] It is a dataset-independent, data-driven process: the age-associated genes are identified empirically within the dataset.

[00376] As a consequence the approach requires 1) enough samples 2) enough age diversity to establish age/ expression levels relationships. The approach is species- and platform-independent and can be applied to any collection metabolite with high-enough dimensionality and dynamic range (for example using proteins by mass-spec, miRNAs by microarray).

[00377] The delta is predicted to reflect the biological age of the studied tissue (by contrast with the chronological age of the organism), thus enabling the systematic spatio-temporal study of aging and its determinants.

[00378] To test the biological relevance of the Delta and the validity of the prediction, 2 main points are tested. First, phenotypic relationships can be tested to determine if samples harboring clinical evidence of premature aging display increased Deltas. For example, in cohorts of samples from individuals affected or not by aging-related pathology (e.g. Alzheimer's Disease in brain, Macular Degeneration in the eye, Pulmonary Fibrosis in the lung), the delta can be calculated for all samples without knowledge of the phenotype and the Delta-Phenotype relationship can be queried. Second, organ specificity can be tested to ask if samples from different tissues from the same organism display different Deltas. Whether deltas relate to aging traits in an organ specific fashion can be queried.

[00379] Example 3 – Identification of genetic trans-modifiers of the phenotypes associated with a genotype of interest

[00380] While some haplotypes, such as those identified by GWAS at TMEM106B, are of clear therapeutic interest, the identity of the gene impacted by the haplotype and how its function is modulated are most often elusive. Even when the identity of the gene and its function are understood, the gene itself might not be easily druggable.

[00381] Without being bound by theory, the impact of a disease/trait-associated haplotype is dependent on other genes that may offer better drug targets. Without being bound by theory, the impact of a disease/trait-associated haplotype may be assessed qualitatively and quantitatively in a hypothesis-free fashion using transcriptome-wide gene expression analysis. Without being bound by theory, variants of interest in such genes would phenocopy the impact of the disease/trait-associated haplotype.

[00382] Described herein is a method of identifying genetic variants modulating TMEM106B genotype impact. The targets of therapeutic interest are variants in druggable

genes at the IL2RA/IL15RA locus phenocopying the effect of the TMEM106B genotype associated with FTD and aging.

[00383] Fig. 29 is a strategy overview for identifying TMEM risk variants. The approach uses a combination of four cohorts with both unaffected and AD individuals: Myers, Harvard, ROSMAP, and Mount Sinai (N~1.5k). The same approach developed for Delta-Age (*see e.g.* Example 1) is applied using genotype instead of age as a variable of interest. The first step in this approach is to identify which genes are modulated by TMEM allele load in unaffected and AD individuals independently. Based on the levels of those genes, the next step is to reverse-predict of TMEM risk allele load within each sample. Within a given TMEM genotype, the next step is to identify which SNP would modulate the apparent TMEM risk allele load. GWAS is then run in each genotype/disease group (Unaffected TMEM RR, Unaffected TMEM PR, Unaffected TMEM PP, AD TMEM RR, AD TMEM PR, AD TMEM PP). Meta-analysis by TMEM genotype is then performed across the disease group. TMEM RR corresponds to an individual homozygous for the risk allele of TMEM; TMEM PR corresponds to an individual heterozygous for the risk/protective allele of TMEM; and TMEM PP corresponds to an individual homozygous for the protective allele of TMEM

[00384] *Results*

[00385] Using the above approach, SNPs at IL2RA were determined to phenocopy TMEM in TMEM RR carriers in a genome wide-significant fashion ($p \sim E-9$) as well as modify Delta Age specifically in TMEM RR. Thus, IL2RA is a genome wide modulator of Delta-TMEM in TMEM RR individuals. IL2RA is associated with Multiple Sclerosis (MS), Rheumatoid Arthritis (RA), Crohn's Disease (CD) and Irritable Bowel Disease (IBD). The risk allele of IL2RA for MS, CD and IBD is associated with lower Delta-Age. The risk allele of IL2RA associated with RA is associated with increased Delta-Age. In addition, a SNP at IL2 is associated with RA. It interacts with IL2RA specifically in the TMEM RR individuals (Interaction $p=9E-3$ in TMEM RR, $p=0.09$ in the same direction in the TMEM PR and nothing in the TMEM PP.)

[00386] Fig. 30 depicts the local association with TMEM phenocopying in TMEM RR individuals at the IL2RA/IL15RA locus.

[00387] Fig. 31 depicts the top hits with LD-based proxies.

[00388] Fig. 32 depicts the effect of IL2RA genotype on Delta-Age in TMEM106B individuals. In the TMEM RR homozygotes, the IL2RA genotype modifies Delta-Aging by up to 20 years. IL2RA is thus shown to be a therapeutic target for the TMEM RR.

[00389] Fig. 33 depicts the effect of TMEM rs1990622 on Delta-Age in the whole cohort, stratified by disease status. Fig. 34 depicts the effect of IL2RA rs12722515 on Delta-Age in the whole cohort, stratified by disease status and TMEM106B genotype. The effect of IL2RA genotype on Delta-Age was highly specific to TMEM106B RR individuals.

[00390] Fig. 35 depicts the cross-sectional rate of cognitive decline measured by Mini-Mental Score Examination, stratified by TMEM 1068 and IL2RA genotypes.

[00391] Fig. 36 depicts the longitudinal rate of temporal atrophy, based on regional MRI measurements at baseline and after 24 months, stratified by TMEM106B and IL2RA genotypes. The effect of IL2RA genotype on Temporal atrophy was highly specific in TMEM106B individuals.

[00392] The modulation of mRNA levels of the identified genes of interest by TMEM genotype, aging and AD was assessed. IL2RA is a component of the trimeric IL2 receptor, together with subunits beta and gamma (IL2RB, IL2RG). IL2RB and IL2RG can also associate with IL15RA to form a trimeric receptor for IL15. The trimeric high affinity IL2 receptor is expressed and functions on cells acquiring an IL-2 signal. Conversely, IL15RA is expressed and binds IL15 with high affinity per se already in the endoplasmic reticulum of the IL15 producing cells and it presents IL15 to cells expressing IL2RB/IL2RG dimeric receptor in trans. Thus, while IL2 is secreted almost exclusively by activated T cells and acts as a free molecule, IL-15 is expressed mostly by myeloid cells and works as a cell surface-associated cytokine.

[00393] Fig. 37 depicts the effect of TMEM risk allele, aging and AD on IL2, IL2RA, IL2RB, IL2RG, IL15 and IL15RA. IL2RB, IL2RG, IL15 and IL15RA show different regulation by TMEM and age.

[00394] Fig. 38 depicts CNS cell type expression pattern of the identified genes of interest and their ligands. IL15RA was strongly upregulated by LPS and not by AD in microglia. IL15 was upregulated in microglia and astrocytes.

[00395] Fig. 39 presents another example of CNS cell type expression pattern of the identified genes of interest and their ligands. IL2RA was upregulated by LPS and decreased by AD in microglia. IL2 was not detected.

[00396] Fig. 40 presents another example of CNS cell type expression pattern of the identified genes of interest and their ligands. IL2RB and IL2RG was upregulated by LPS and decreased by AD in microglia.

[00397] Fig. 41 depicts a table of genetic modifiers of TMEM106B and the top hits using GWAS.

[00398] Fig. 42 shows genes which show a pattern of expression similar to IL15RA in the LPS dataset. Genes include, but are not limited to, CCL2, TLR2, PILRB, TREM1, TMEM106A, TSPC.

[00399] Example 4 - Genetic determinants of pathways of interest

[00400] The functional annotation of gene expression studies is conducted using gene sets based on gene expression patterns (multiple variables, corresponding to a large number of genes). Described herein are efficient tools of genome-wide association studies to query the association between millions of genotypes and a single variable (either disease status or a quantitative trait) across the genotyped samples.

[00401] Without being bound by theory, genes in a given pathway are likely to be co-regulated. Coordinated variation in a set of genes involved in a pathway of interest is an indicator of the activation of this pathway. Genetic variants in the genome may modify the activity of key regulators of pathways of interest.

[00402] Fig. 43 shows a strategy overview. All the 285 genes from the Synaptic Membrane GO gene set were grouped into 1 meta-gene representing the average levels of all its members to visualize directly the effect of TMEM106B genotype on aggregated synaptic genes levels. As shown in Figs. 44-46, the cellular compartment GO category the most decreased by TMEM106B risk allele in human brain is Synaptic Membrane.

[00403] Fig. 47 shows genome-wide scan for genetic determinants of Synaptic genes levels in human brain. TMEM106B is the main genetic determinant of synaptic genes levels in human brain.

[00404] Figs. 48-49 show the effect of TMEM106B genotype on aggregated Synaptic genes levels in human brain. Fig. 48 shows the effect of TMEM106B genotype and disease status on aggregated synaptic genes levels. TMEM106B risk allele is associated with less synapses in disease-free individuals, but also in AD or HD patients. Fig. 49 shows the effect of TMEM106B genotype and age on aggregated synaptic genes levels in unaffected. TMEM106B risk allele is associated with increased age-associated rate of synaptic loss in neurodegenerative-free individuals.

[00405] Fig. 50 shows the effect of TMEM106B genotype on specific synaptic genes from the gene set.

[00406] Figs. 51-53 show that the cellular compartment GO category the most increased by TMEM106B risk allele in human brain is Lysosomal Lumen.

What is claimed is:

1. A computer-implemented method of determining a biological age of a sample from a subject comprising:
 - a) providing a gene expression level of a plurality of genes in a sample from a subject;
 - b) determining, for each gene whose expression level is significantly correlated with chronological age, a residual value, wherein the residual value is determined by comparing the gene expression level of said gene in the sample from the subject to a linear regression of the expression level of the gene as a function of chronological age for a plurality of individuals;
 - c) determining, for each gene whose expression level is significantly correlated with chronological age, a ratio of the residual value determined in step (b) to the linear regression coefficient for the gene;
 - d) repeating steps (b) and (c) for each gene whose expression level is significantly correlated with chronological age; and
 - e) integrating the ratios for each gene whose expression level is significantly correlated with chronological age, wherein the integrated ratio corresponds to the biological age of the sample from the subject.
2. A computer-implemented method of determining a biological age of a sample from a subject comprising:
 - a) providing a gene expression level of a plurality of genes in a sample for a plurality of individuals;
 - b) performing a linear regression of the expression level of each gene of the plurality of genes as a function of chronological age of each of the plurality of individuals to determine a linear regression co-efficient for each gene of the plurality of genes;
 - c) selecting, from the plurality of genes, the genes whose expression level is significantly correlated with chronological age;
 - d) determining, for each gene whose expression level is significantly correlated with chronological age selected in step (c), a residual value, wherein the residual value is determined by comparing the gene expression level of said gene in the sample from the subject to the linear regression performed in step (b) for said gene;

- e) determining, for each gene whose expression level is significantly correlated with chronological age, a ratio of the residual value determined in step (d) to the linear regression coefficient determined in step (b);
 - f) repeating steps (d) and (e) for each gene whose expression level is significantly correlated with chronological age selected in step (c); and
 - g) integrating the ratios for each gene whose expression level is significantly correlated with chronological age, wherein the integrated ratio corresponds to the biological age of the sample from the subject.
3. The method of claim 1 or 2, further comprising: comparing the biological age to a chronological age of the subject to determine a differential aging value for the subject.
 4. The method of claim 1, 2, or 3, wherein the plurality of individuals includes the subject.
 5. The method of claim 1, wherein steps (c) to (e) are performed for each individual in the plurality of individuals.
 6. The method of claim 2, wherein steps (d) to (g) are performed for each individual in the plurality of individuals.
 7. The method of claim 1 to 6, further comprising performing a genome-wide association study (GWAS).
 8. The method of claim 7, wherein the GWAS identifies single-nucleotide polymorphism (SNP) modifiers of the differential aging value in the plurality of individuals.
 9. The method of claim 7, wherein the GWAS identifies genetic modifiers of the differential aging value in the plurality of individuals.
 10. The method of claims 1 to 9, wherein the gene expression level is the gene expression level in the brain.
 11. The method of claim 10, wherein the gene expression level is the gene expression level in the frontal cortex.
 12. The method of claim 10, wherein the gene expression level is the gene expression level in the cerebellum.

13. A computer program product for determining a biological age of a sample from a subject comprising instructions which, when the program is executed by a computer, cause the computer to carry out the steps of:
- a) providing a gene expression level of a plurality of genes in a sample for a plurality of individuals;
 - b) determining, for each gene whose expression level is significantly correlated with chronological age, a residual value, wherein the residual value is determined by comparing the gene expression level of said gene in the sample from the subject to a linear regression of the expression level of the gene as a function of chronological age for a plurality of individuals;
 - c) determining, for each gene whose expression level is significantly correlated with chronological age, a ratio of the residual value determined in step (b) to the linear regression coefficient for the gene;
 - d) repeating steps (b) and (c) for each gene whose expression level is significantly correlated with chronological age; and
 - e) integrating the ratios for each gene whose expression level is significantly correlated with chronological age, wherein the integrated ratio corresponds to the biological age of the sample from the subject.
14. A computer program product for determining a biological age of a sample from a subject comprising instructions which, when the program is executed by a computer, cause the computer to carry out the steps of:
- a) providing a gene expression level of a plurality of genes in a sample for a plurality of individuals;
 - b) performing a linear regression of the expression level of each gene of the plurality of genes as a function of chronological age of each of the plurality of individuals to determine a linear regression co-efficient for each gene of the plurality of genes;
 - c) selecting, from the plurality of genes, the genes whose expression level is significantly correlated with chronological age;
 - d) determining, for each gene whose expression level is significantly correlated with chronological age selected in step (c), a residual value, wherein the residual value is determined by comparing the gene expression level of said gene in the sample from the subject to the linear regression performed in step (b) for said gene;

- e) determining, for each gene whose expression level is significantly correlated with chronological age, a ratio of the residual value determined in step (d) to the linear regression coefficient determined in step (b);
- f) repeating steps (d) and (e) for each gene whose expression level is significantly correlated with chronological age selected in step (c); and
- g) integrating the ratios for each gene whose expression level is significantly correlated with chronological age, wherein the integrated ratio corresponds to the biological age of the sample from the subject.

15. The computer program product of claim 13 or 14, further comprising carrying out the step of: comparing the biological age to a chronological age of the subject to determine a differential aging value for the subject.

16. The computer program product of claim 13, 14, or 15, wherein the plurality of individuals includes the subject.

17. The computer program product of claim 13, wherein steps (c) to (e) are performed for each individual in the plurality of individuals.

18. The computer program product of claim 14, wherein steps (d) to (g) are performed for each individual in the plurality of individuals.

19. The computer program product of claim 13 to 18, further comprising carrying out the step of performing a genome-wide association study (GWAS).

20. The computer program product of claim 19, wherein the GWAS identifies single-nucleotide polymorphism (SNP) modifiers of the differential aging value in the plurality of individuals.

21. The computer program product of claim 19, wherein the GWAS identifies genetic modifiers of the differential aging value in the plurality of individuals.

22. The computer program product of claims 13 to 21, wherein the gene expression level is the gene expression level in the brain.

23. The computer program product of claim 22, wherein the gene expression level is the gene expression level in the frontal cortex.

24. The computer program product of claim 22, wherein the gene expression level is the gene expression level in the cerebellum.

25. A computer-implemented method of determining a phenotype of a sample from a subject, wherein the phenotype is correlated with a haplotype of interest, the method comprising:

- a) providing a gene expression level of a plurality of genes in a sample for a plurality of individuals, wherein each individual's haplotype of interest is known;
- b) performing a linear regression of the expression level of each gene of the plurality of genes as a function of the haplotype of interest for each of the plurality of individuals to determine a linear regression co-efficient for each gene of the plurality of genes;
- c) selecting, from the plurality of genes, the genes whose expression level is significantly correlated with the haplotype of interest;
- d) determining, for each gene whose expression level is significantly correlated with the haplotype of interest selected in step (c), a residual value, wherein the residual value is determined by comparing the gene expression level of said gene in a sample from a subject to the linear regression performed in step (b) for said gene;
- e) determining, for each gene whose expression level is significantly correlated with the haplotype of interest, a ratio of the residual value determined in step (d) to the linear regression coefficient determined in step (b);
- f) repeating steps (d) and (e) for each gene whose expression level is significantly correlated with the haplotype of interest selected in step (c); and
- g) integrating the ratios for each gene whose expression level is significantly correlated with the haplotype of interest, wherein the integrated ratio corresponds to the phenotype of the sample from the subject.

26. The method of claim 25, wherein the plurality of individuals includes the subject.

27. The method of claim 25, wherein steps (d) to (g) are performed for each individual in the plurality of individuals.

28. The method of claims 25 to 27, further comprising: h) performing a genome-wide association study (GWAS).

29. The method of claim 28, wherein the GWAS identifies single-nucleotide polymorphism (SNP) modifiers of the phenotype correlated with the haplotype of interest in the plurality of individuals.
30. The method of claim 28, wherein the GWAS identifies genetic modifiers of the phenotype in the plurality of individuals.
31. The method of claim 25 to 30, wherein the haplotype of interest is defined as 0, 1, or 2 allele copies.
32. The method of claim 31, wherein the allele copies are determined by SNP genotyping.
33. The method of claim 25 to 32, wherein the phenotype correlated with the haplotype of interest is an expression level of a plurality of genes.
34. The method of claims 25 to 33, wherein the gene expression level is the gene expression level in the brain.
35. The method of claim 34, wherein the gene expression level is the gene expression level in the frontal cortex.
36. The method of claim 34, wherein the gene expression level is the gene expression level in the cerebellum.
37. A computer program product for determining the phenotype of a sample from a subject, wherein the phenotype is correlated with a haplotype of interest, comprising instructions which, when the program is executed by a computer, cause the computer to carry out the steps of:
- a) providing a gene expression level of a plurality of genes in a sample for a plurality of individuals, wherein each individual's haplotype of interest is known;
 - b) performing a linear regression of the expression level of each gene of the plurality of genes as a function of the haplotype of interest for each of the plurality of individuals to determine a linear regression co-efficient for each gene of the plurality of genes;

- c) selecting, from the plurality of genes, the genes whose expression level is significantly correlated with the haplotype of interest;
- d) determining, for each gene whose expression level is significantly correlated with the haplotype of interest selected in step (c), a residual value, wherein the residual value is determined by comparing the gene expression level of said gene in a sample from a subject to the linear regression performed in step (b) for said gene;
- e) determining, for each gene whose expression level is significantly correlated with the haplotype of interest, a ratio of the residual value determined in step (d) to the linear regression coefficient determined in step (b);
- f) repeating steps (d) and (e) for each gene whose expression level is significantly correlated with the haplotype of interest selected in step (c); and
- g) integrating the ratios for each gene whose expression level is significantly correlated with the haplotype of interest, wherein the integrated ratio corresponds to the phenotype of the sample from the subject.

- 38. The computer program product of claim 37, wherein the plurality of individuals includes the subject.
- 39. The computer program product of claim 37, wherein steps (d) to (g) are performed for each individual in the plurality of individuals.
- 40. The computer program product of claims 37 to 39, further comprising carrying out the step of: h) performing a genome-wide association study (GWAS).
- 41. The computer program product of claim 40, wherein the GWAS identifies single-nucleotide polymorphism (SNP) modifiers of the phenotype in the plurality of individuals.
- 42. The computer program product of claim 40, wherein the GWAS identifies genetic modifiers of the phenotype in the plurality of individuals.
- 43. The computer program product of claims 37 to 42, wherein the haplotype of interest is defined as 0, 1, or 2 allele copies.
- 44. The computer program product of claim 43, wherein the allele copies are determined by SNP genotyping.

45. The computer program product of claims 37 to 44, wherein the phenotype correlated with a haplotype of interest is an expression level of a plurality of genes.
46. The computer program product of claims 37 to 45, wherein the gene expression level is the gene expression level in the brain.
47. The computer program product of claim 46, wherein the gene expression level is the gene expression level in the frontal cortex.
48. The computer program product of claim 46, wherein the gene expression level is the gene expression level in the cerebellum.
49. A computer-implemented method of identifying one or more genetic modifiers of a plurality of genes whose expression level is correlated with a genotype of interest comprising:
- a) providing a gene expression level of a plurality of genes in a sample for a plurality of individuals, wherein each individual's genotype of interest is known;
 - b) performing a linear regression of the expression level of each gene of the plurality of genes as a function of the genotype of interest of each of the plurality of individuals to determine a linear regression co-efficient for each gene of the plurality of genes;
 - c) selecting, from the plurality of genes, the genes whose expression level is significantly correlated with the genotype of interest;
 - d) determining, for each gene whose expression level is significantly correlated with the genotype of interest selected in step (c), a residual value, wherein the residual value is determined by comparing the gene expression level of said gene in a sample from a subject to the linear regression performed in step (b) for said gene;
 - e) determining, for each gene whose expression level is significantly correlated with the genotype of interest, a ratio of the residual value determined in step (d) to the linear regression coefficient determined in step (b);
 - f) repeating steps (d) and (e) for each gene whose expression level is significantly correlated with the genotype of interest selected in step (c);
 - g) integrating the ratios for each gene significantly correlated with the genotype of interest;
 - h) performing steps (d) to (g) for each individual in the plurality of individuals; and

i) identifying genetic modifiers that modulate the integrated ratio value determined in step (g).

50. The method of claim 49, wherein the genetic modifiers that modulate the integrated ratio value determined in step (g) are determined by performing a genome-wide association study (GWAS).

51. The method of claim 49, wherein the plurality of individuals includes the subject.

52. The method of claim 50, wherein the GWAS identifies single-nucleotide polymorphism (SNP) modifiers of the integrated ratio value determined in step (g) in the plurality of individuals.

53. The method of claim 49, wherein the genetic modifier is a single-nucleotide polymorphism (SNP).

54. The method of claims 49 to 53, wherein the genotype of interest is the risk allele of TMEM106B.

55. The method of claim 54, wherein the risk allele of TMEM106B is and A at SNP rs1990622.

56. The method of claims 49 to 53, wherein the genotype of interest is a risk allele associated with a disease or disorder.

57. The method of claims 49 to 53, wherein the genotype of interest is a non-risk allele.

58. The method of claims 49 to 53, wherein the genotype of interest is a haplotype.

59. The method of claims 49 to 53, wherein the gene expression level is the gene expression level in the brain.

60. The method of claims 49 to 53, wherein the gene expression level is the gene expression level in the frontal cortex.

61. The method of claims 49 to 53, wherein the gene expression level is the gene expression level in the cerebellum.

62. The method of claims 49 to 53, wherein the plurality of individuals are healthy individuals.
63. The method of claims 49 to 53, wherein the plurality of individuals have a disease or disorder.
64. The method of claims 49 to 53, wherein the plurality of individuals have a neurodegenerative disease.
65. The method of claim 64, wherein the neurodegenerative disease is Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, or Fronto-temporal dementia.
66. A computer program product for identifying one or more genetic modifiers of a plurality of genes whose expression level is correlated with a genotype of interest comprising instructions which, when the program is executed by a computer, cause the computer to carry out the steps of:
- a) providing a gene expression level of a plurality of genes in a sample for a plurality of individuals, wherein each individual's genotype of interest is known;
 - b) performing a linear regression of the expression level of each gene of the plurality of genes as a function of the genotype of interest of each of the plurality of individuals to determine a linear regression co-efficient for each gene of the plurality of genes;
 - c) selecting, from the plurality of genes, the genes whose expression level is significantly correlated with the genotype of interest;
 - d) determining, for each gene whose expression level is significantly correlated with the genotype of interest selected in step (c), a residual value, wherein the residual value is determined by comparing the gene expression level of said gene in a sample from a subject to the linear regression performed in step (b) for said gene;
 - e) determining, for each gene whose expression level is significantly correlated with the genotype of interest, a ratio of the residual value determined in step (d) to the linear regression coefficient determined in step (b);
 - f) repeating steps (d) and (e) for each gene whose expression level is significantly correlated with the genotype of interest selected in step (c);
 - g) integrating the ratios for each gene significantly correlated with the genotype of interest;
 - h) performing steps (d) to (g) for each individual in the plurality of individuals; and

i) identifying genetic modifiers that modulate the integrated ratio value determined in step (g).

67. The computer program product of claim 66, wherein the genetic modifiers that modulate the integrated ratio value determined in step (g) are determined by performing a genome-wide association study (GWAS).

68. The computer program product of claim 66, wherein the plurality of individuals includes the subject.

69. The computer program product of claim 67, wherein the GWAS identifies single-nucleotide polymorphism (SNP) modifiers of the integrated ratio value determined in step (g) in the plurality of individuals.

70. The computer program product of claim 66, wherein the genetic modifier is a single-nucleotide polymorphism (SNP).

71. The computer program product of claims 66 to 70, wherein the genotype of interest is the risk allele of TMEM106B.

72. The computer program product of claim 71, wherein the risk allele of TMEM106B is and A at SNP rs1990622.

73. The computer program product of claims 66 to 70, wherein the genotype of interest is a risk allele associated with a disease or disorder.

74. The computer program product of claims 66 to 70, wherein the genotype of interest is a non-risk allele.

75. The computer program product of claims 66 to 70, wherein the genotype of interest is a haplotype.

76. The computer program product of claims 66 to 70, wherein the gene expression level is the gene expression level in the brain.

77. The computer program product of claims 66 to 70, wherein the gene expression level is the gene expression level in the frontal cortex.

78. The computer program product of claims 66 to 70, wherein the gene expression level is the gene expression level in the cerebellum.
79. The computer program product of claims 66 to 70, wherein the plurality of individuals are healthy individuals.
80. The computer program product of claims 66 to 70, wherein the plurality of individuals have a disease or disorder.
81. The computer program product of claims 66 to 70, wherein the plurality of individuals have a neurodegenerative disease.
82. The method of claim 81, wherein the neurodegenerative disease is Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, or Fronto-temporal dementia.
83. A method of modifying a phenotype associated with a TMEM106B risk allele in a subject in need thereof, the method comprising administering an effective amount of an IL2 modulator, and IL2RA modulator, an IL2RB modulator, an IL2RG modulator, an IL15 modulator, an IL15RA modulator, or a combination thereof to the subject.
84. A method of treating, preventing, or delaying the onset of aging in a subject in need thereof, the method comprising administering an effective amount of an IL2 modulator, and IL2RA modulator, an IL2RB modulator, an IL2RG modulator, an IL15 modulator, an IL15RA modulator, a TMEM106B modulator, or a combination thereof to the subject.
85. A method of treating, preventing, or delaying the onset of cognitive decline in a subject in need thereof, the method comprising administering an effective amount of an IL2 modulator, and IL2RA modulator, an IL2RB modulator, an IL2RG modulator, an IL15 modulator, an IL15RA modulator, a TMEM106B modulator, or a combination thereof to the subject.
86. The method of claims 83-85, wherein the subject is administered an IL2 modulator.
87. The method of claims 83-85, wherein the subject is administered an IL2RA modulator.
88. The method of claims 83-85, wherein the subject is administered an IL2RB modulator.

89. The method of claims 83-85, wherein the subject is administered an IL2RG modulator.
90. The method of claims 83-85, wherein the subject is administered an IL15 modulator.
91. The method of claims 83-85, wherein the subject is administered an IL15RA modulator.
92. The method of claims 83-85, wherein the subject is homozygous for the TMEM106B risk allele.
93. The method of claims 83-85, wherein the subject is heterozygous for the TMEM106B risk allele.
94. The method of claim 83-85, wherein the subject is homozygous for the TMEM106B protective allele.
95. The method of claims 92-93, wherein the TMEM106B risk allele is an A at SNP rs1990622.
96. The method of claim 94, wherein the TMEM106B protective allele is a G at SNP rs1990622.
97. The method of claims 83-85, wherein the IL2RA modulator increases expression of a IL2RA protective allele, or decreases expression of a IL2RA risk allele, or a combination thereof.
98. The method of claim 97, wherein the IL2RA protective allele is an A at SNP rs12722515.
99. The method of claim 97, wherein the IL2RA risk allele is an C at SNP rs12722515.
100. The method of claims 83-85, wherein the modulation increases expression of a TMEM106B protective allele.
101. The method of claim 83-85, wherein the modulation decreases the expression of the TMEM106B risk allele.

102. The method of claims 101, wherein the TMEM106B risk allele is an A at SNP rs1990622.
103. The method of claim 100, wherein the TMEM106B protective allele is a G at SNP rs1990622.
104. The method of claim 83, wherein the phenotype associated with a TMEM106B risk allele is a plurality of genes, and their expression levels, associated with the TMEM106B risk allele.
105. The method of claim 86, wherein the phenotype associated with a TMEM106B risk allele is reduced and a phenotype associated with a TMEM106B protective allele is increased.
106. The method of claim 83 wherein the phenotype associated with a TMEM106B risk allele is a plurality of genes whose expression level is correlated with the TMEM106B risk allele.

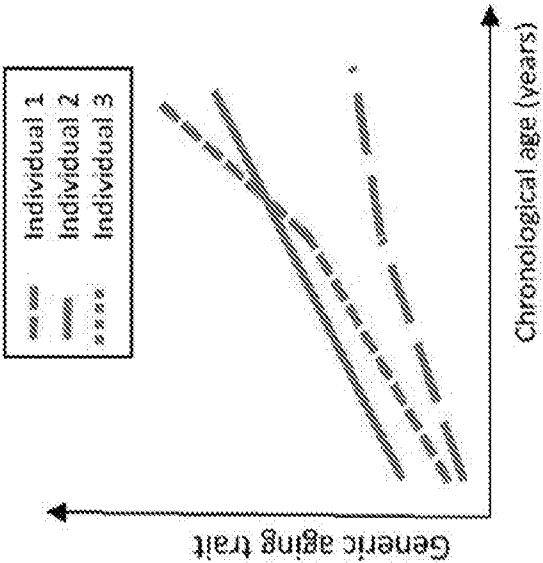


Fig. 1A

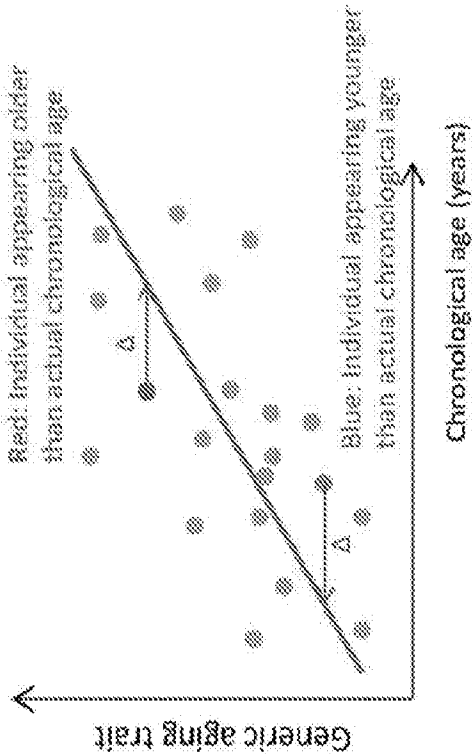


Fig. 1B

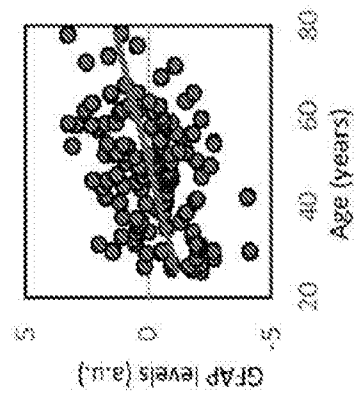


Fig. 2C

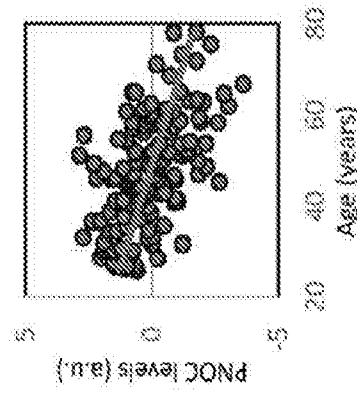


Fig. 2B

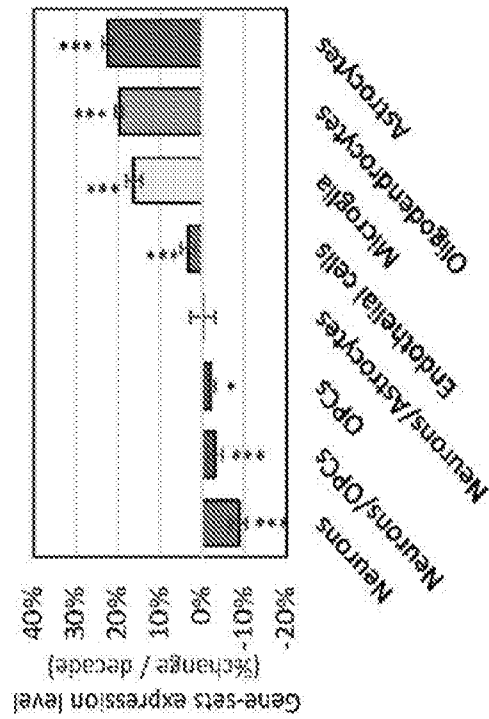


Fig. 2A

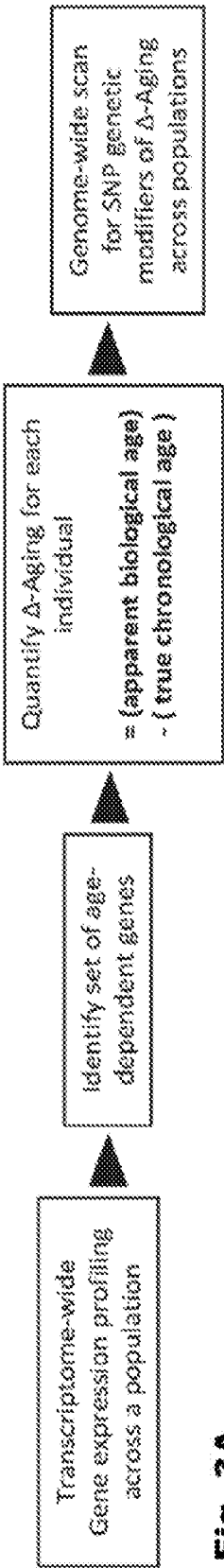


Fig. 3A

Stage	Dataset Name	Status	in all adults (>25yo)			in younger only (<65yo)			in older only (>65yo)		
			Effect (yrs/allele)	n	p	Effect (yrs/allele)	n	p	Effect (yrs/allele)	n	p
Discovery	Tgm/HSTBC/BrainEq/BrainCloud	Und/Und/Und/Und	-4.50	716	2.77x10 ⁻⁷	-1.52	303	0.08	-6.69	413	5.41x10 ⁻¹⁰
Replication	RedMap	Long	-5.54	497	1.68x10 ⁻¹¹	NA	0	NA	-5.54	497	1.68x10 ⁻¹¹
Discovery +Replication			-4.52	1213	1.50x10 ⁻¹⁰	-1.52	303	0.08	-6.06	910	2.56x10 ⁻¹²
Diseases	Tgm/HSTBC/HSTBC	AD/AD/SD	-3.72	601	1.29x10 ⁻⁷	1.04	145	0.77	-4.75	546	9.51x10 ⁻⁶
All combined			-4.69	1303	1.48x10 ⁻¹⁰	-0.76	448	0.85	-5.57	1456	1.20x10 ⁻¹⁰

Fig. 3B

Fig. 3D

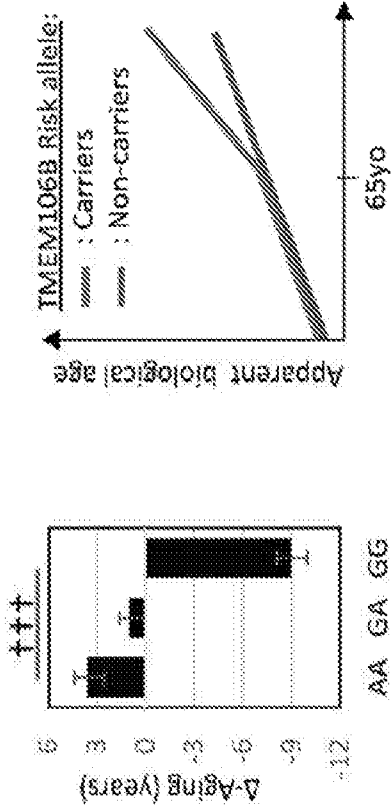
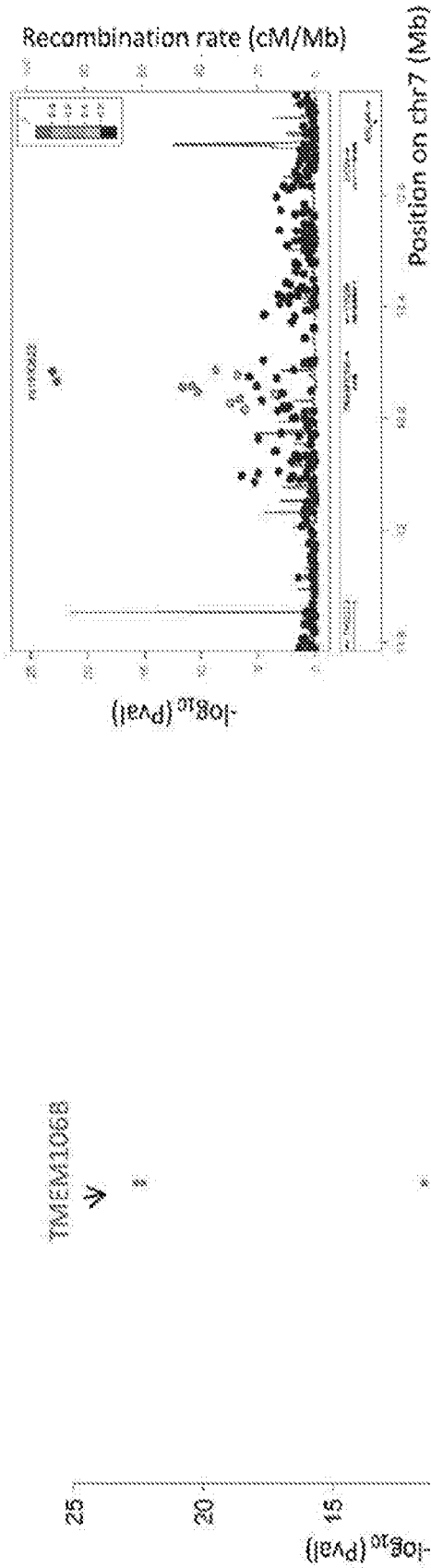


Fig. 3E

Fig. 3F

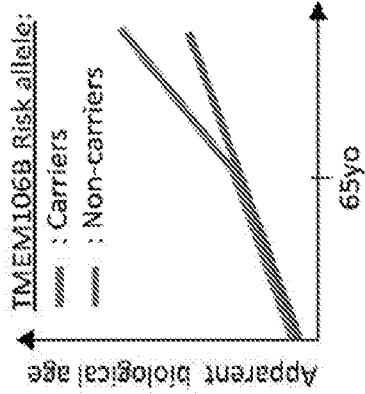


Fig. 3C

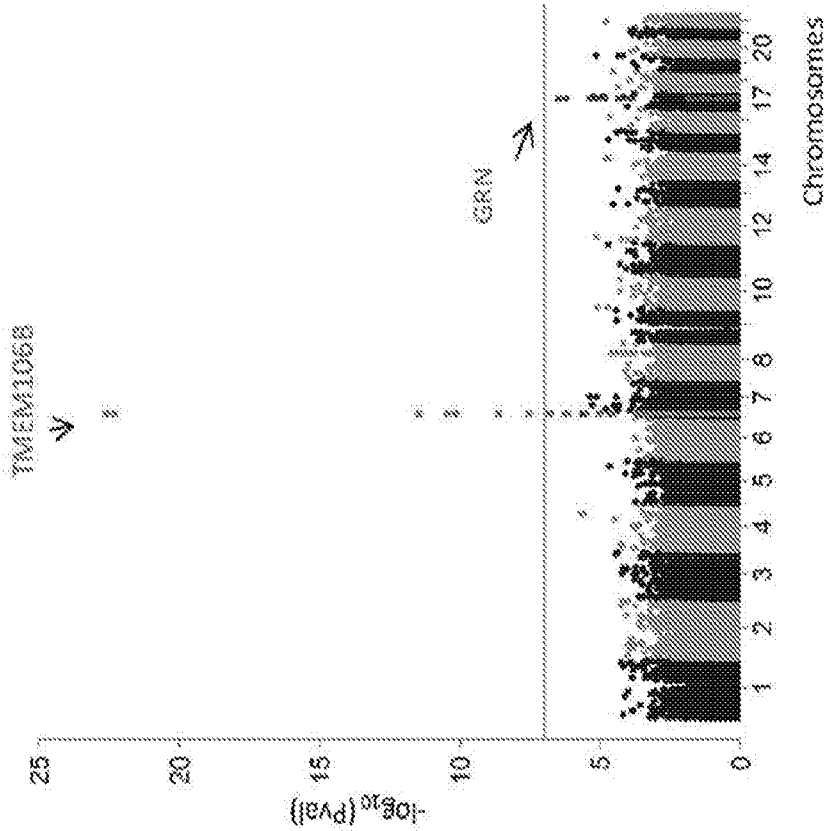


Fig. 4B

Factor	MMSE , LongLife Study			
	< 65yo Individuals (n=1731)		> 65yo Individuals (n=2154)	
	Effect	p	Effect	p
TMEM106B Risk allele load	Age	-0.01/year	0.13	-0.17/year 2.2x10 ⁻³⁶
		-0.03/G allele	0.68	0.26/G allele 7.7x10 ⁻³

Fig. 4C

Factor	Memory Score , Health and Retirement Study			
	50-65yo Individuals (n=6605)		65-80yo Individuals (n=5606)	
	Effect	p	Effect	p
TMEM106B Risk allele load	Age	-0.08/year 1.2x10 ⁻³³	-0.16/year 9.8x10 ⁻³⁸	
		0.007/G allele 0.88	0.13/ G allele 2.7x10 ⁻²	

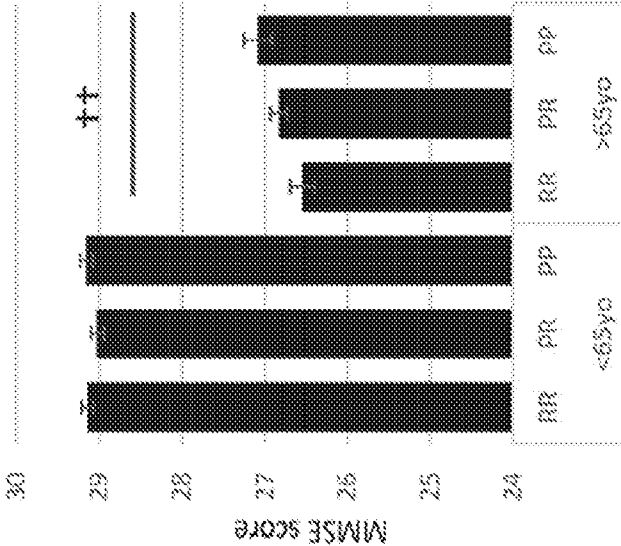


Fig. 4A

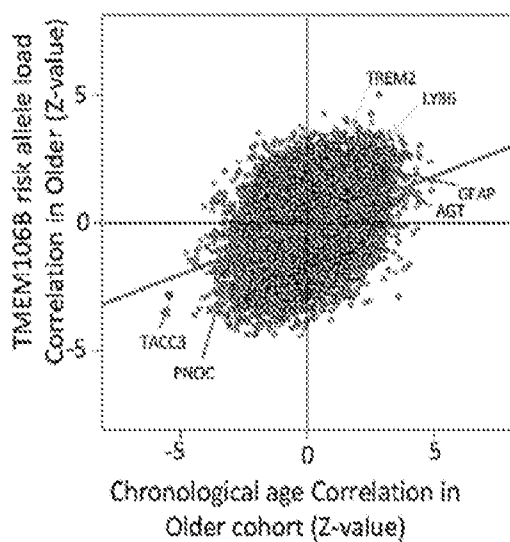
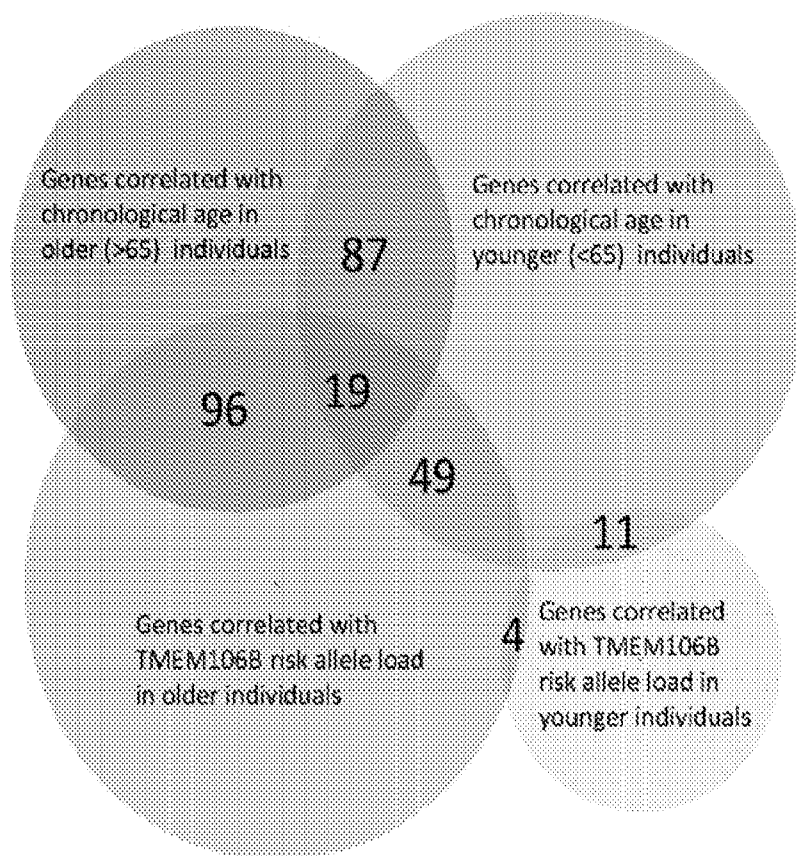
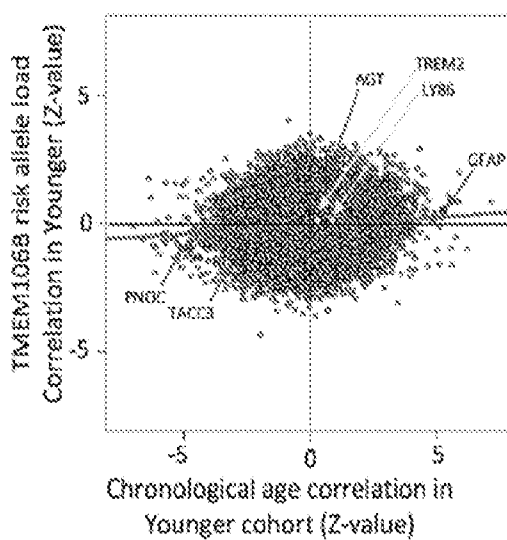
Fig. 5A**Fig. 5B****Fig. 5C**

Fig. 6A

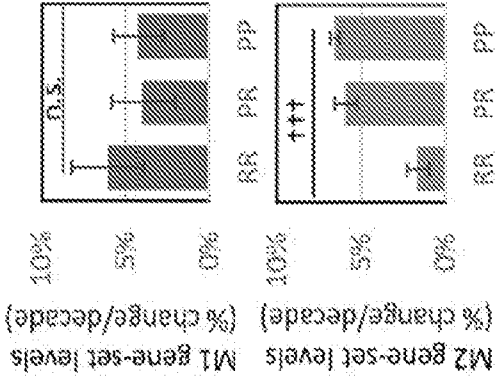
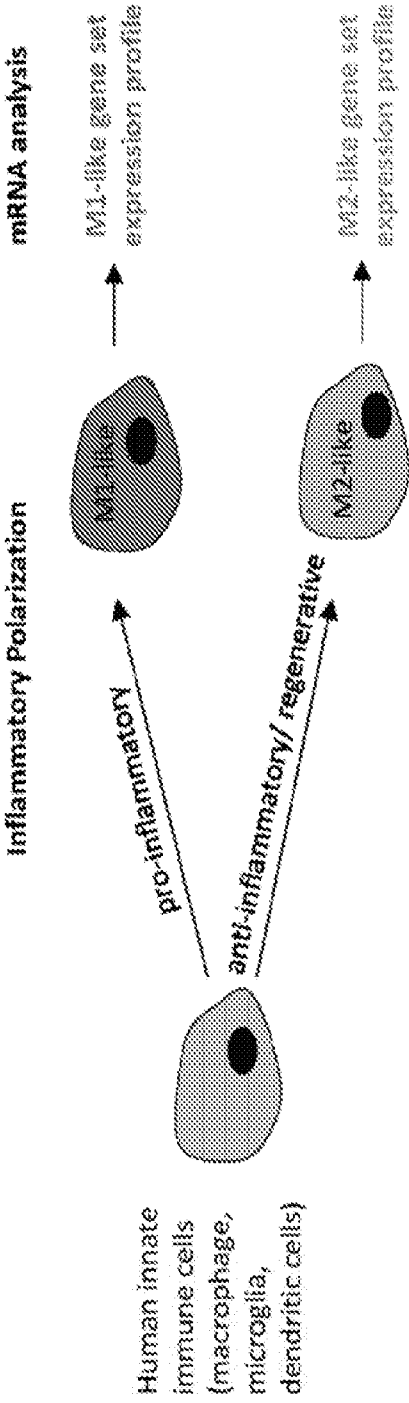


Fig. 6B

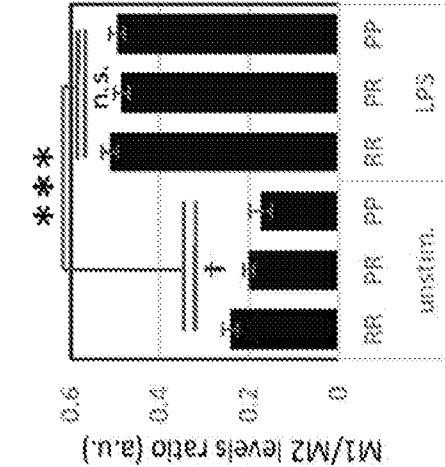


Fig. 6C

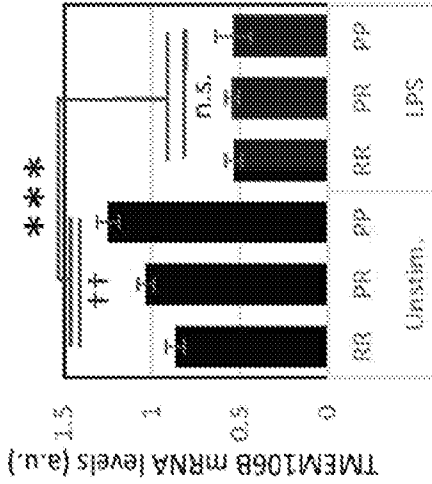


Fig. 6D

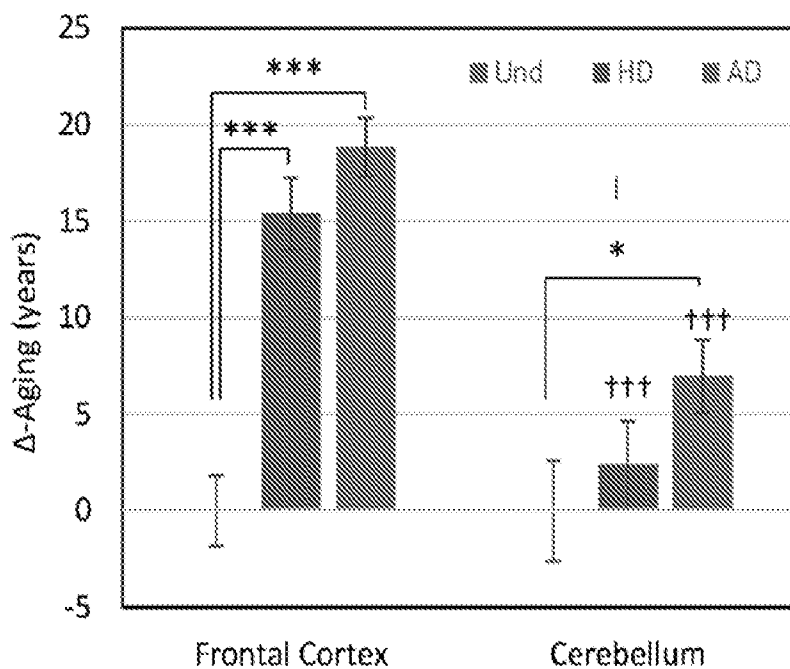
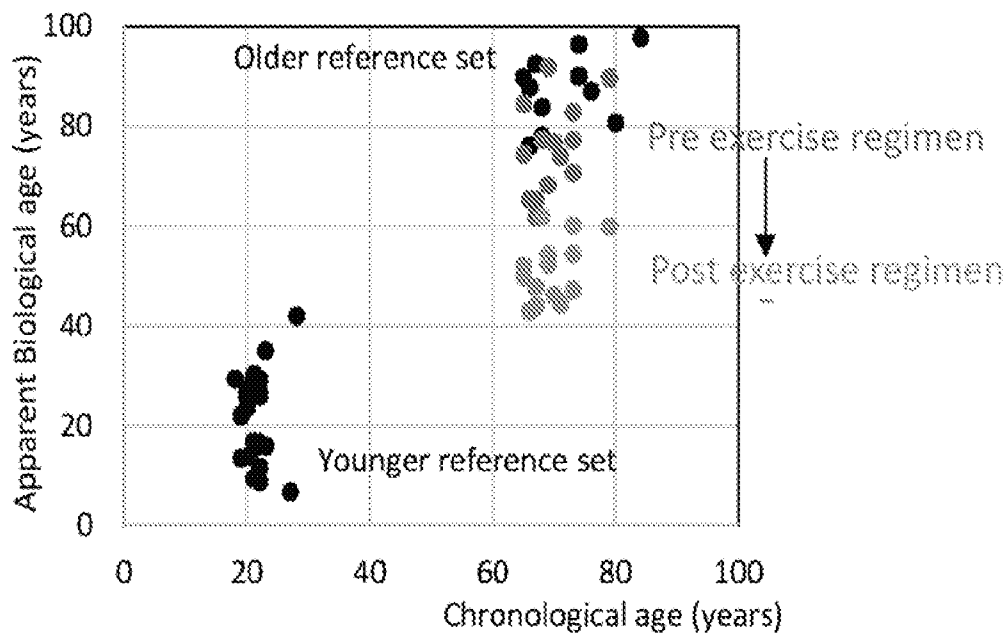
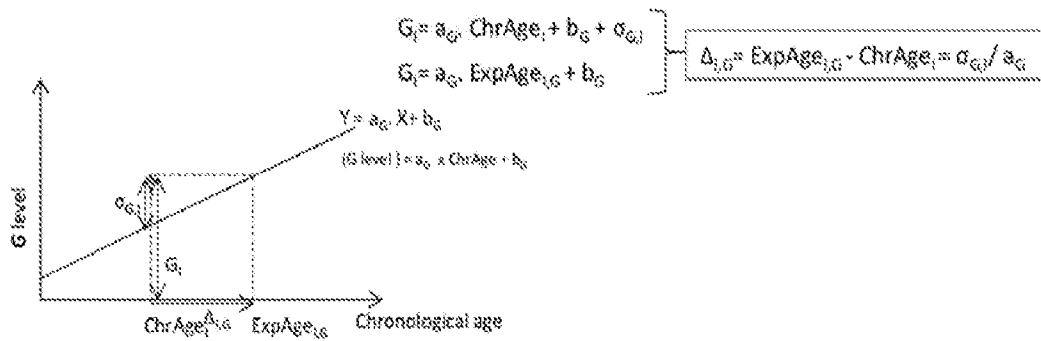
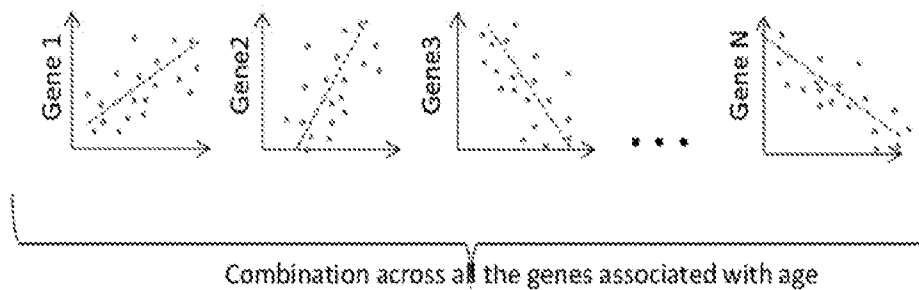
Fig. 7A**Fig. 7B**

Fig. 8A**Fig. 8B**

for a given individual by integrating all the genes affected by aging:

$$\Delta_i = \frac{1}{N} \sum_{G=1}^N \frac{\sigma_{G,i}}{a_G}$$

Delta Age for individual i

Residual for individual i of a linear fit of G levels in function of age across individuals

Linear regression coefficient of a linear fit of G levels in function of age across individuals

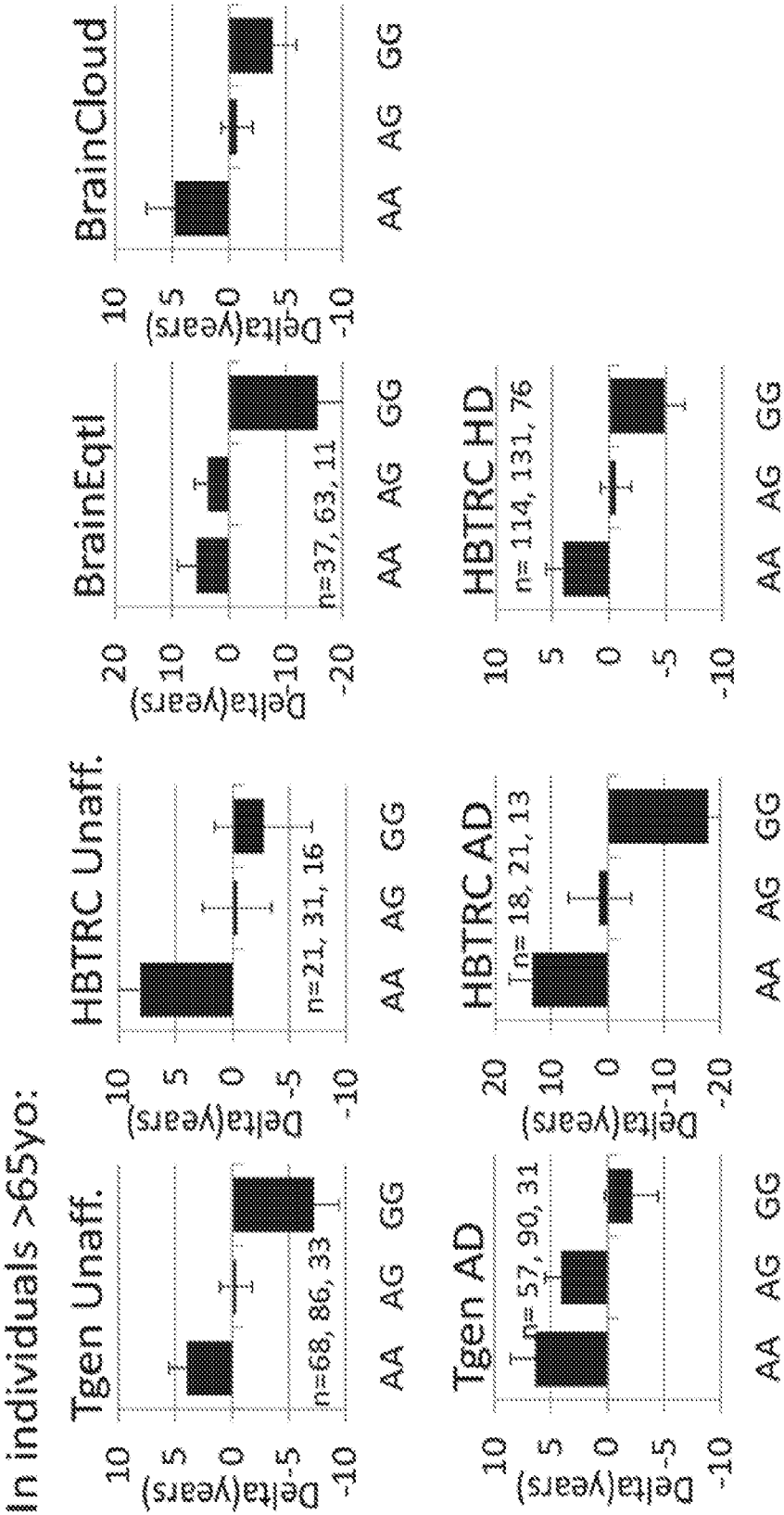


Fig. 9

In individuals <65yo:

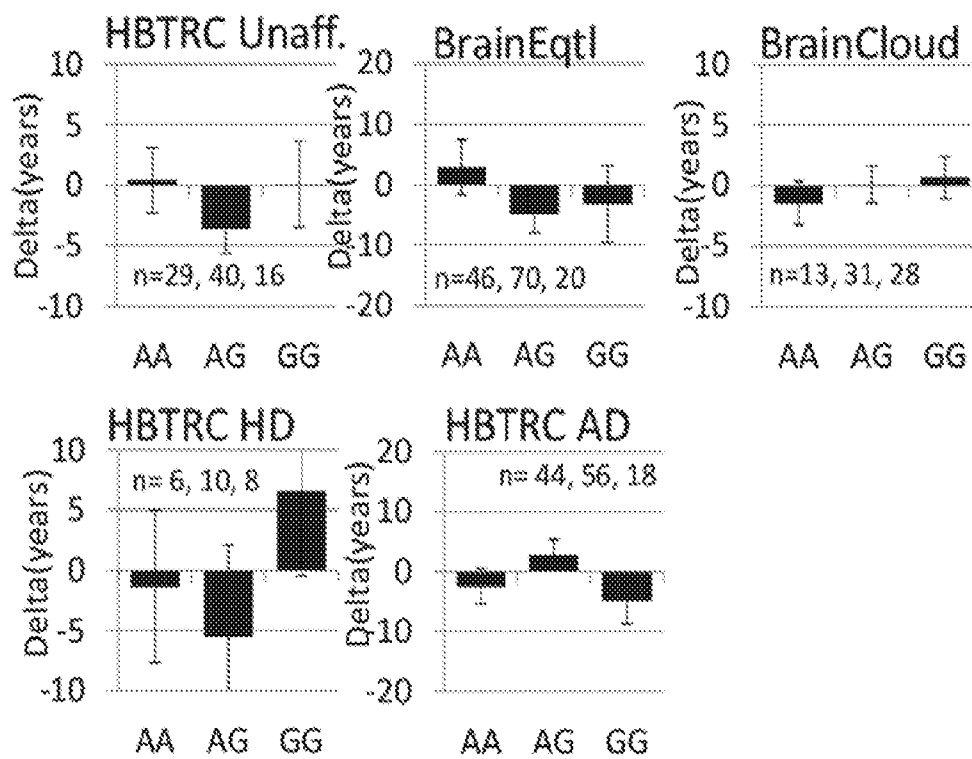
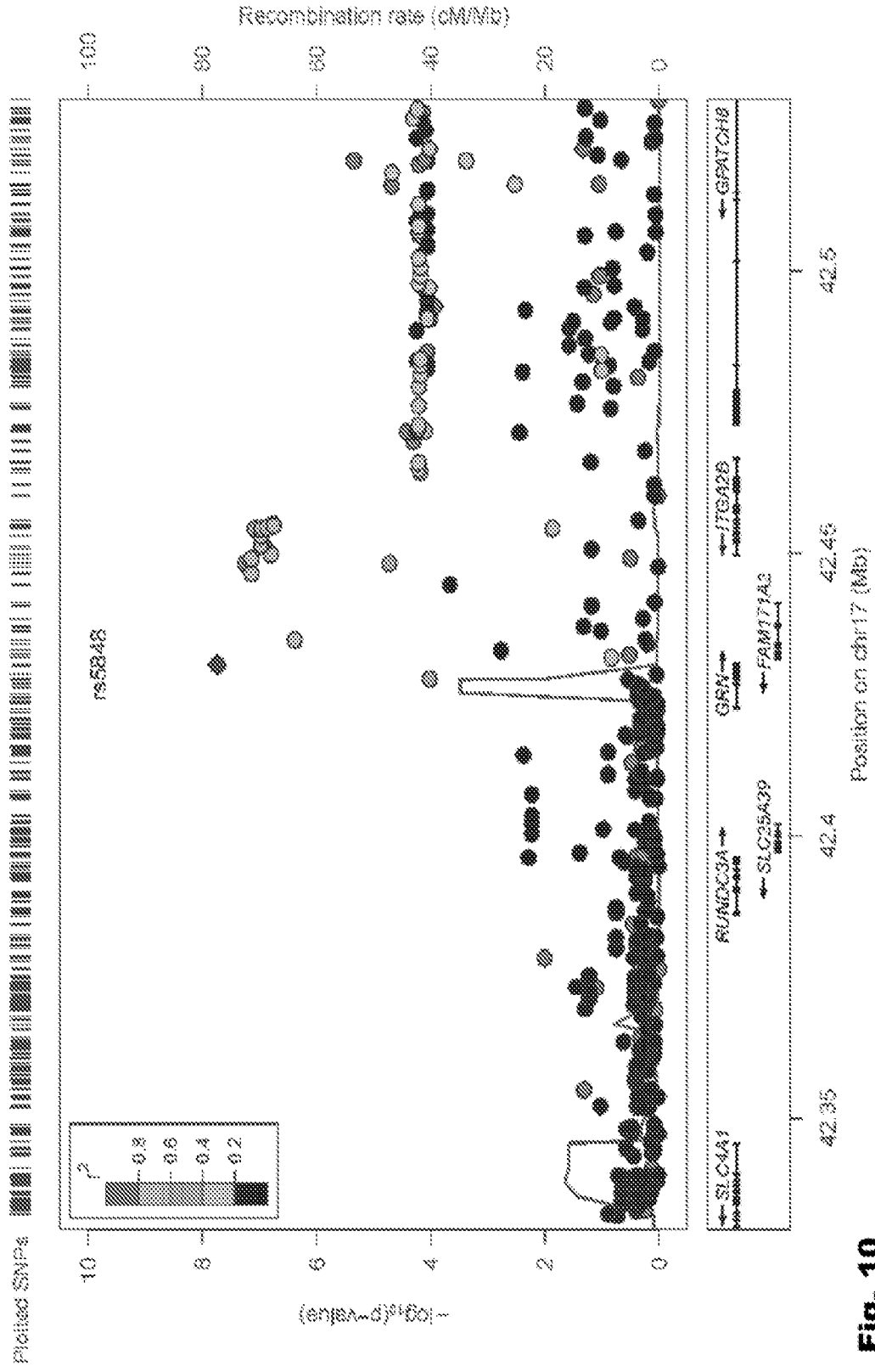


Fig. 9 cont.



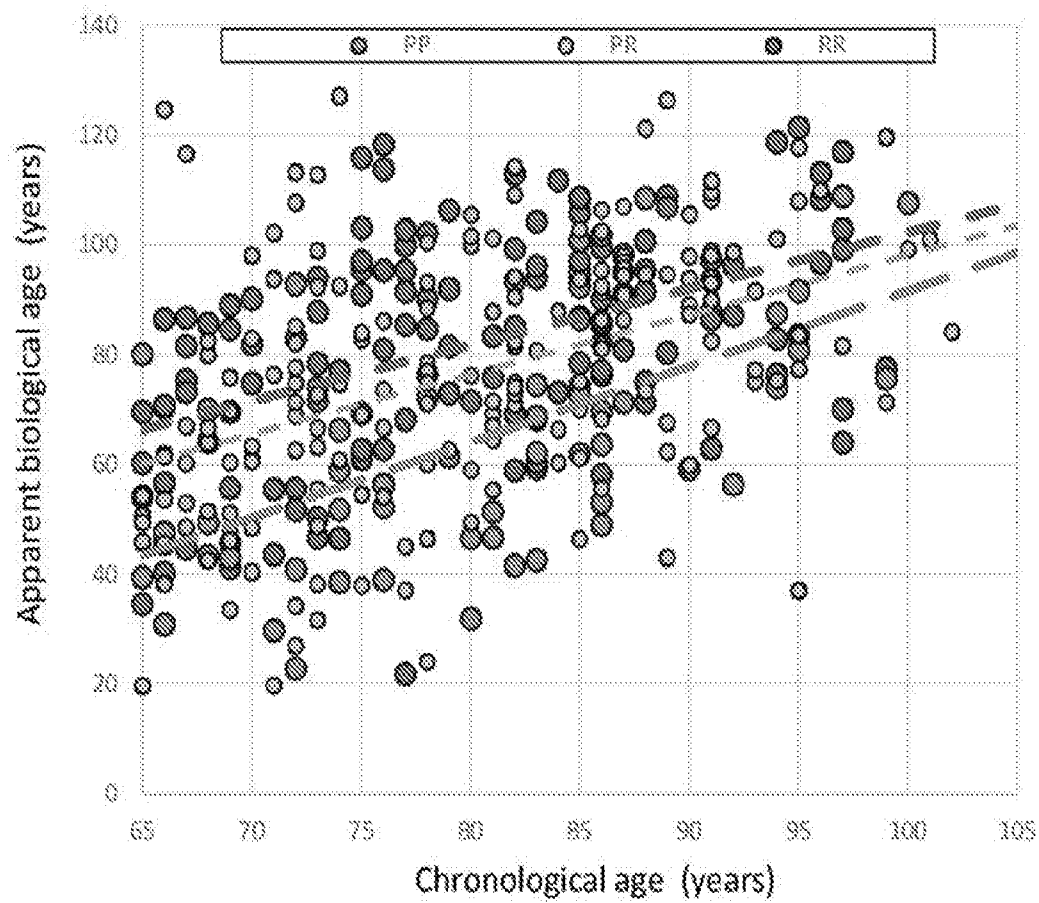
**Fig. 11**

Fig. 12A

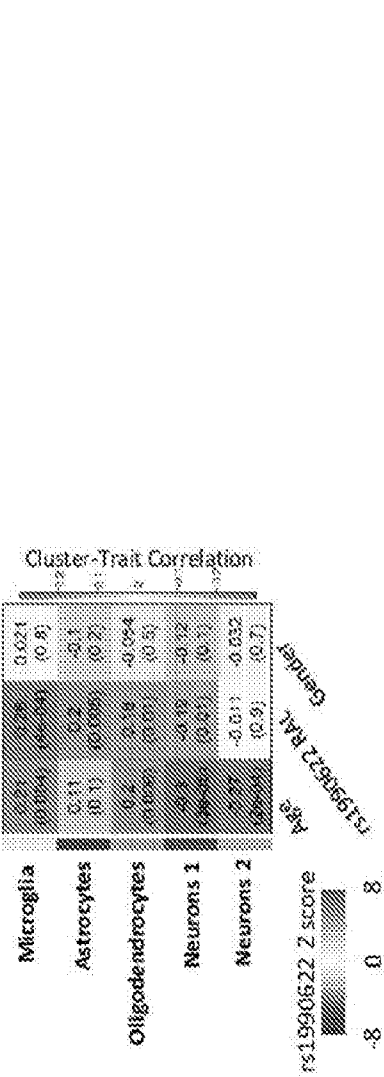
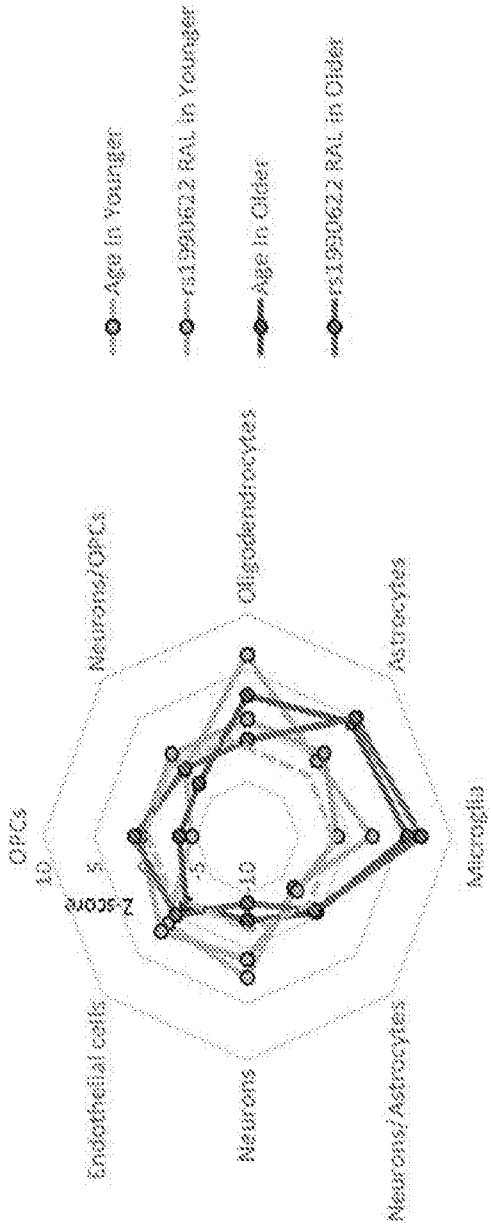


Fig. 12B



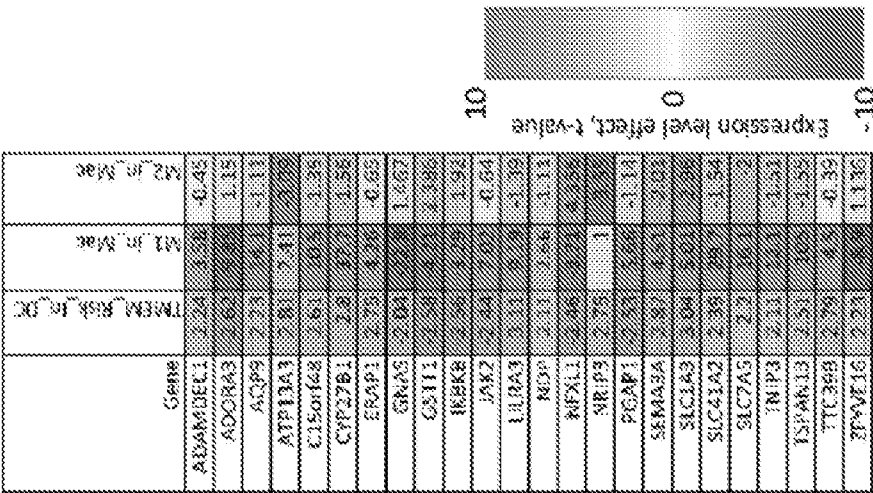


Fig. 13A

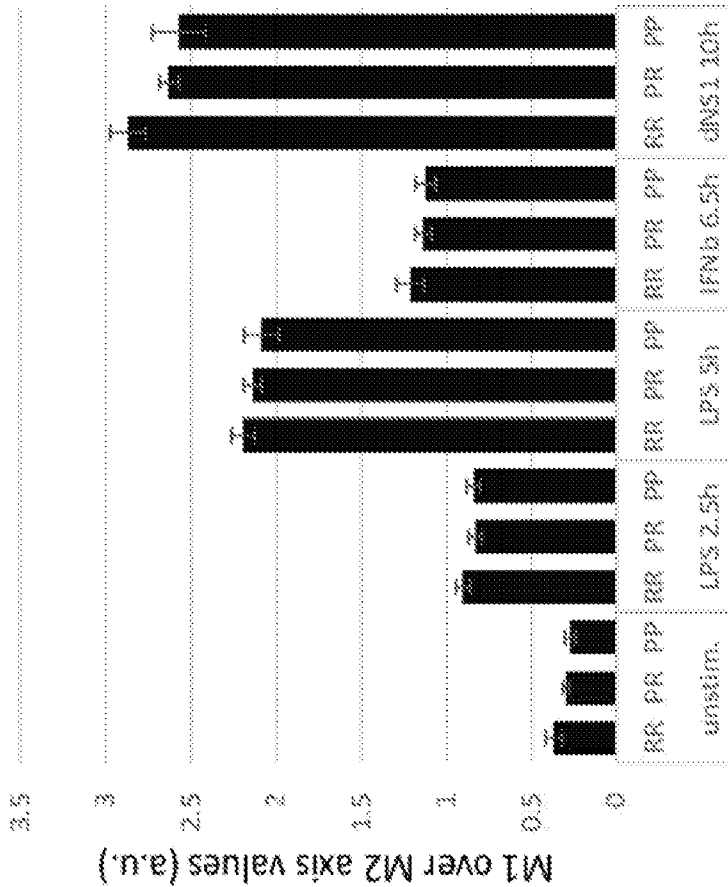


Fig. 13B

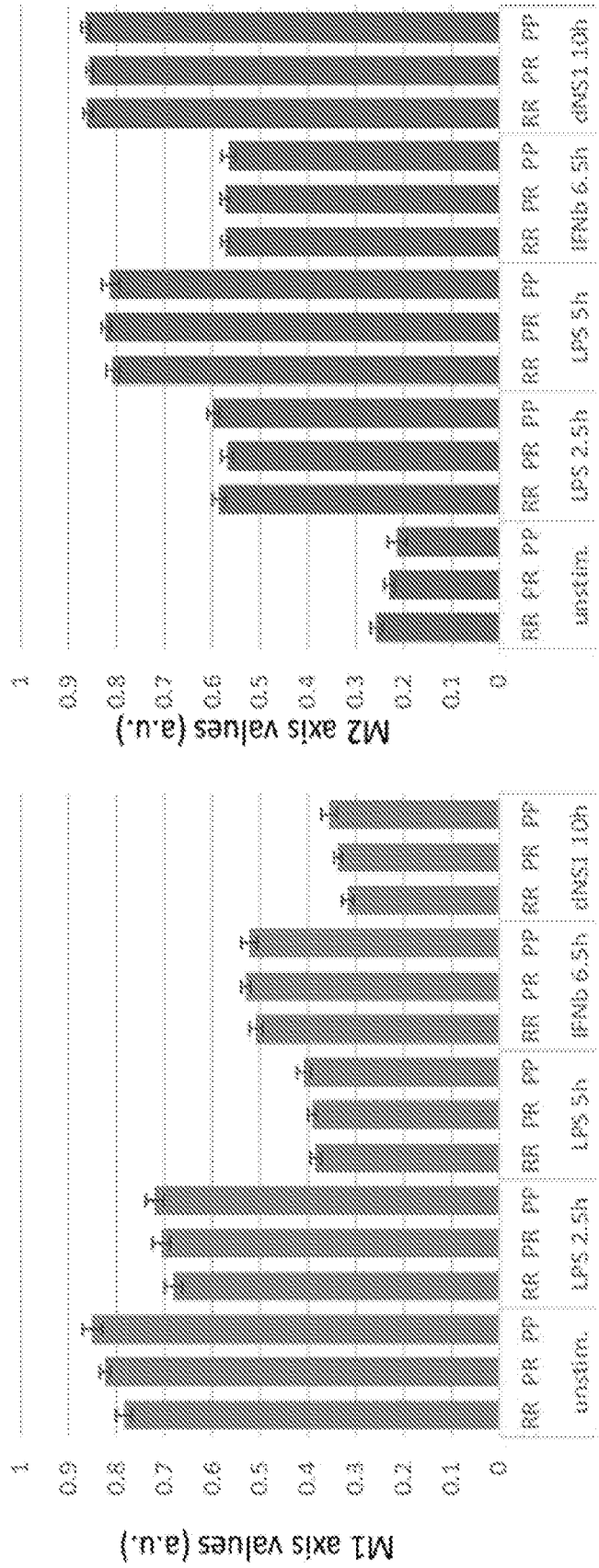


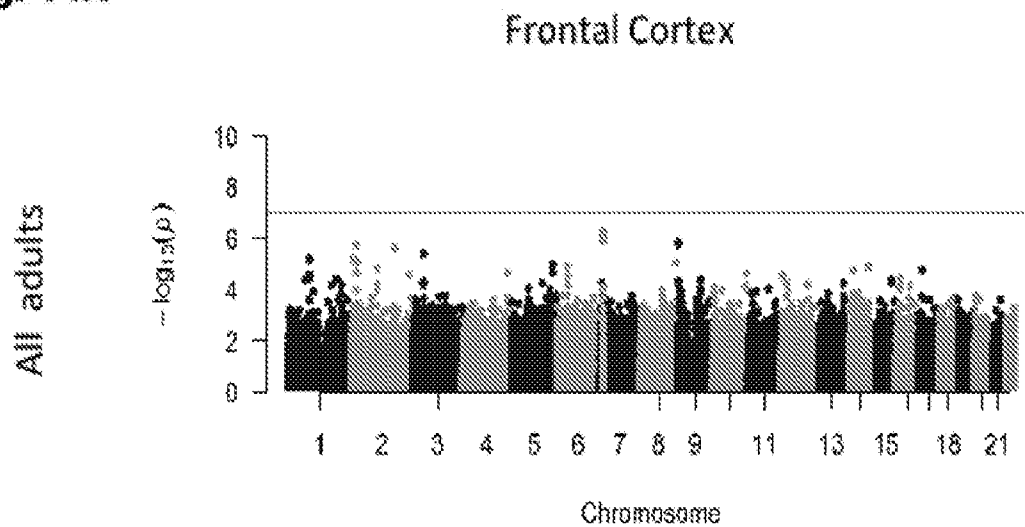
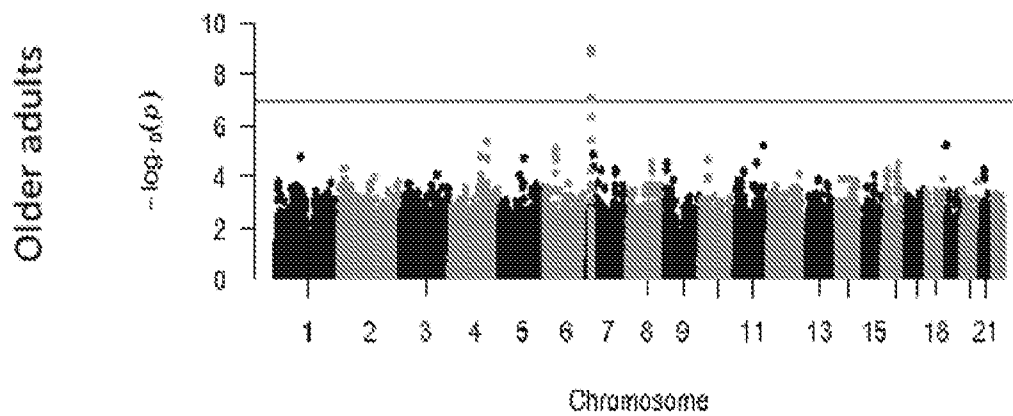
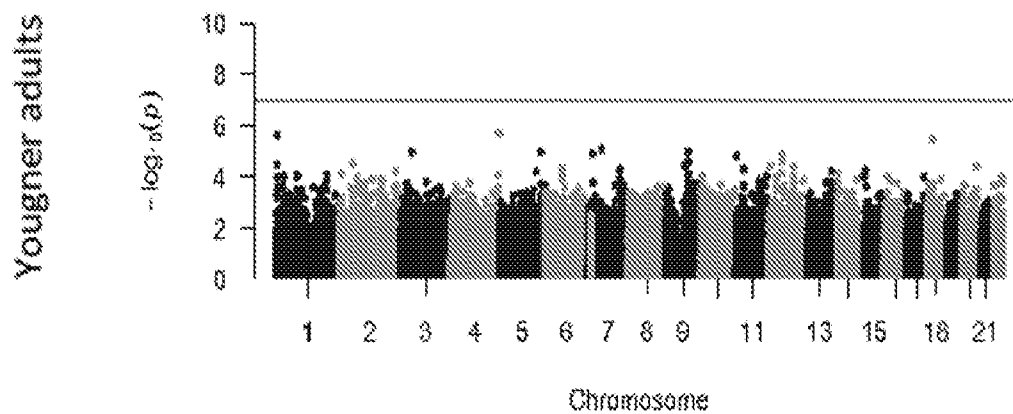
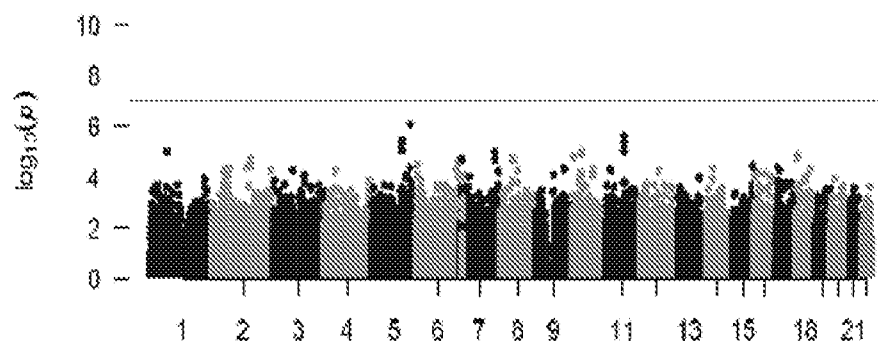
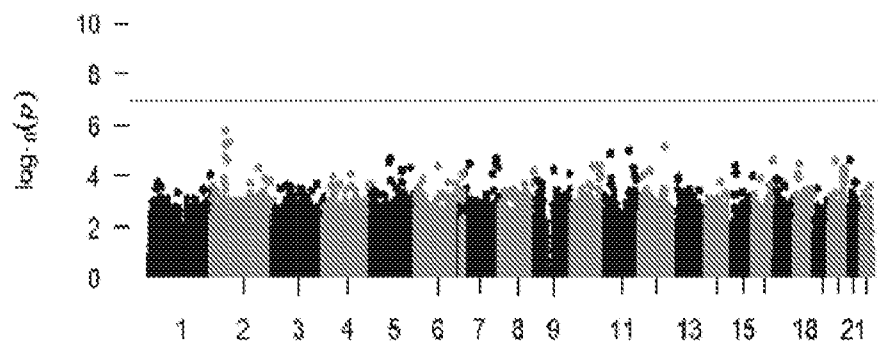
Fig. 14A**Fig. 14B****Fig. 14C**

Fig. 14D

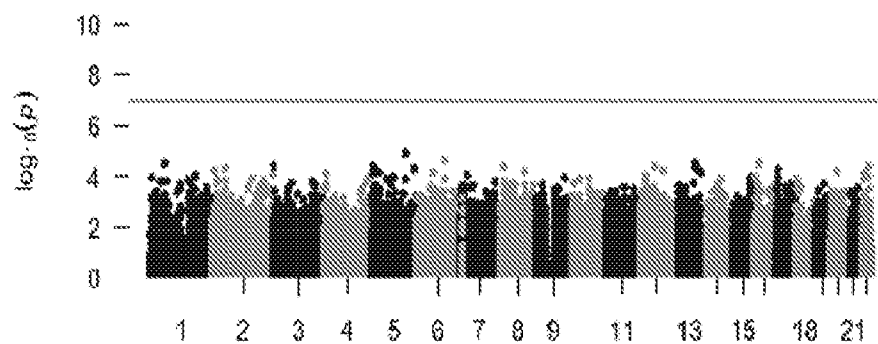
Cerebellum

**Fig. 14E**

Chromosome

**Fig. 14F**

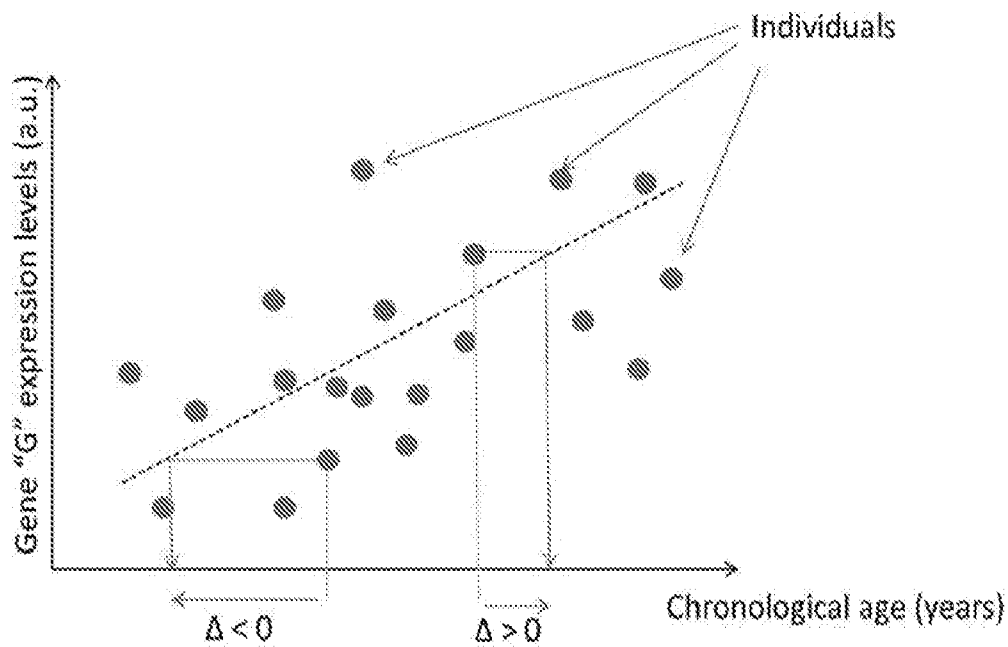
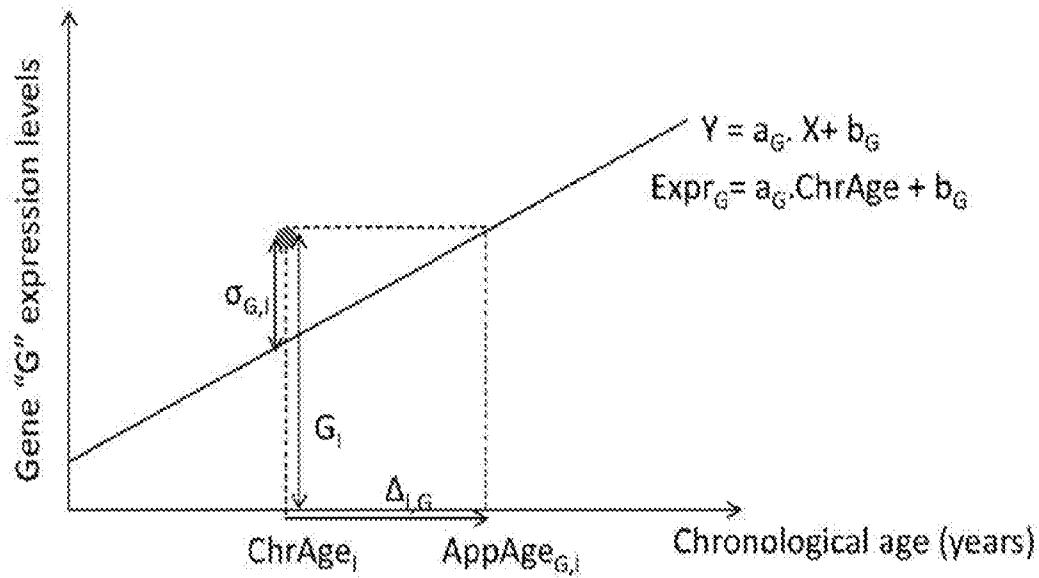
Chromosome



Chromosome

Stage	Dataset Name	References	Status	In all adults (>25yo)			In younger only (<65yo)			In older only (>65yo)		
				Effect {yrs/allele}	n	p	Effect {yrs/allele}	n	p	Effect {yrs/allele}	n	p
Discovery	Tgen	(Myers et al., 2007; Webster et al., 2009)	Und	-6.43	179	8.36E-05	NA	0	NA	-6.43	179	8.36E-05
Discovery	H6TRC	(Zhang et al., 2013)	Und	-2.82	153	8.97E-02	-0.79	91	8.13E-01	-6.51	62	1.56E-02
Discovery	BrainTest	(Gibbs et al., 2010)	Und	-5.31	256	1.93E-02	-3.17	140	3.33E-01	-8.03	116	8.57E-03
Discovery	BrainCloud	(Colanthoni et al., 2011)	Und	-2.17	128	1.98E-02	-3.01	72	9.91E-01	-4.94	56	4.75E-04
Replication	RosMap	(Chan et al., 2015)	Long	-5.54	497	1.68E-15	NA	0	NA	-5.54	497	1.68E-15
Disease	Tgen	(Myers et al., 2007; Webster et al., 2009)	AD	-2.08	176	1.10E-01	NA	0	NA	-2.08	176	1.10E-01
Disease	H6TRC	(Zhang et al., 2013)	AD	-3.84	345	8.82E-04	4.47	25	4.23E-01	-4.47	320	1.45E-04
Disease	H6TRC	(Zhang et al., 2013)	HD	-5.17	179	2.22E-02	0.32	120	8.84E-01	-15.93	50	2.12E-04

Fig. 15

**Fig. 16****Fig. 17**

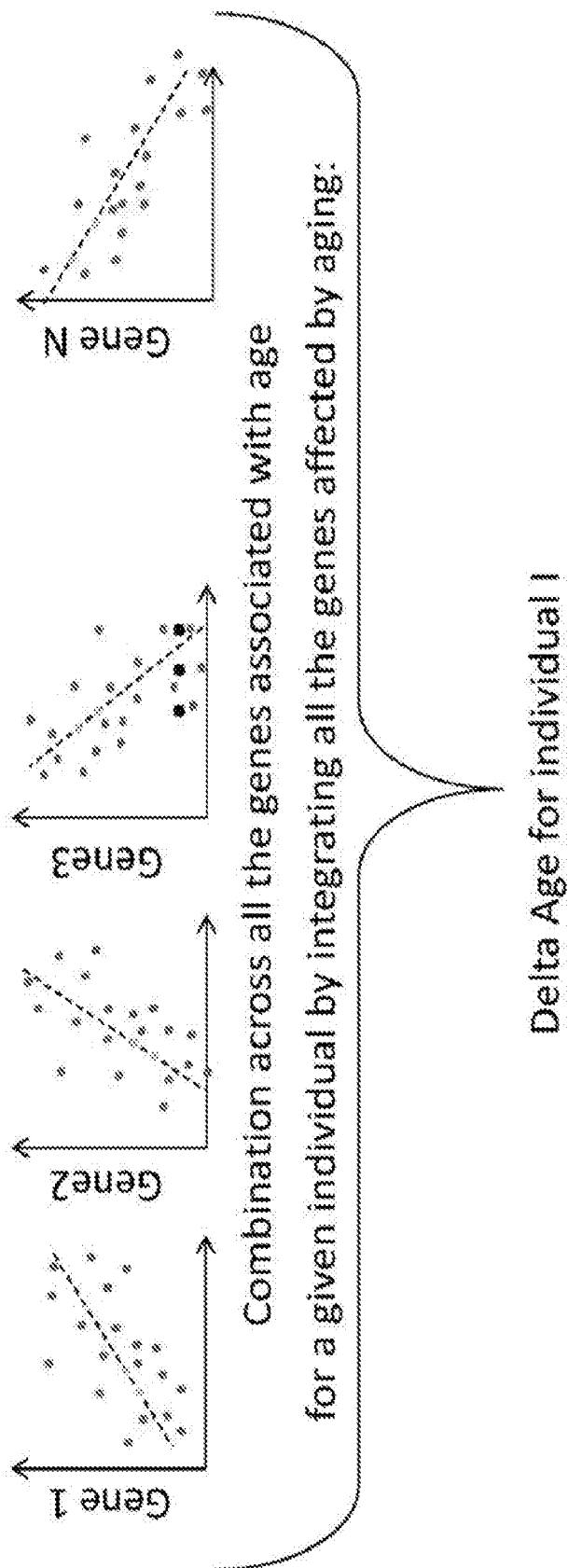


Fig. 18

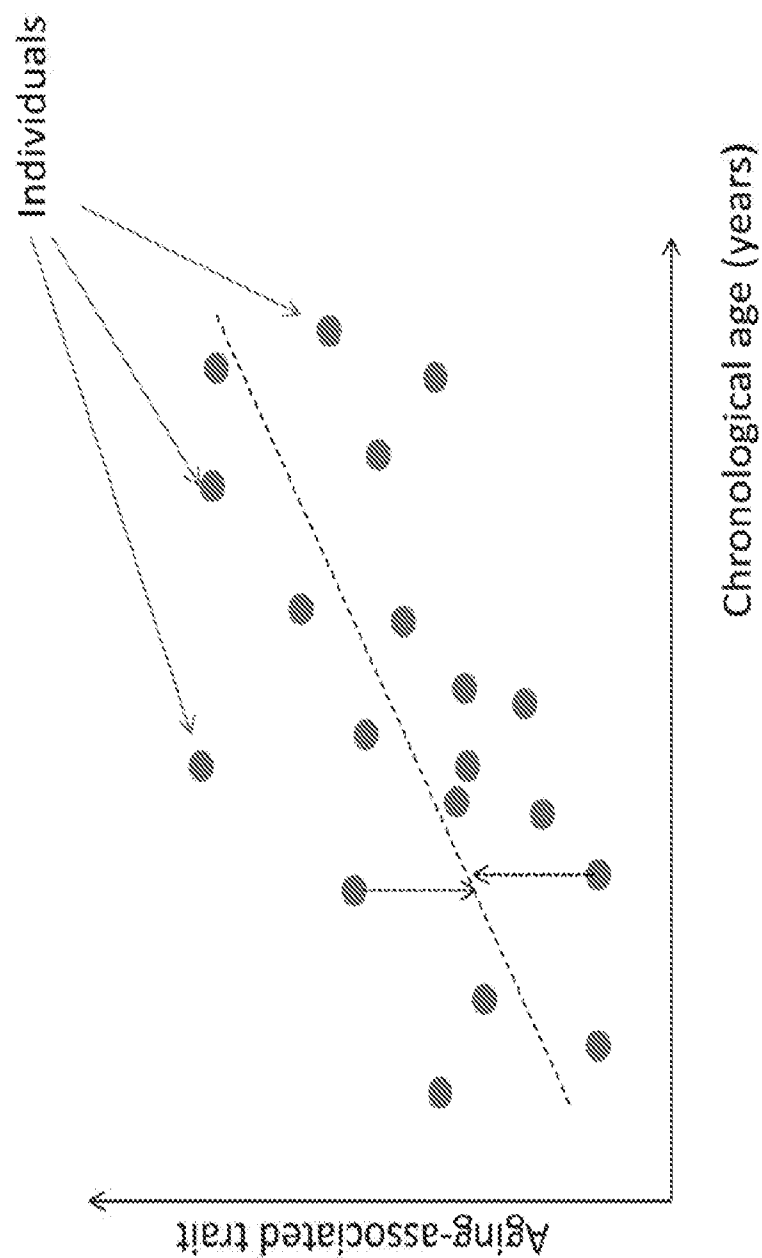
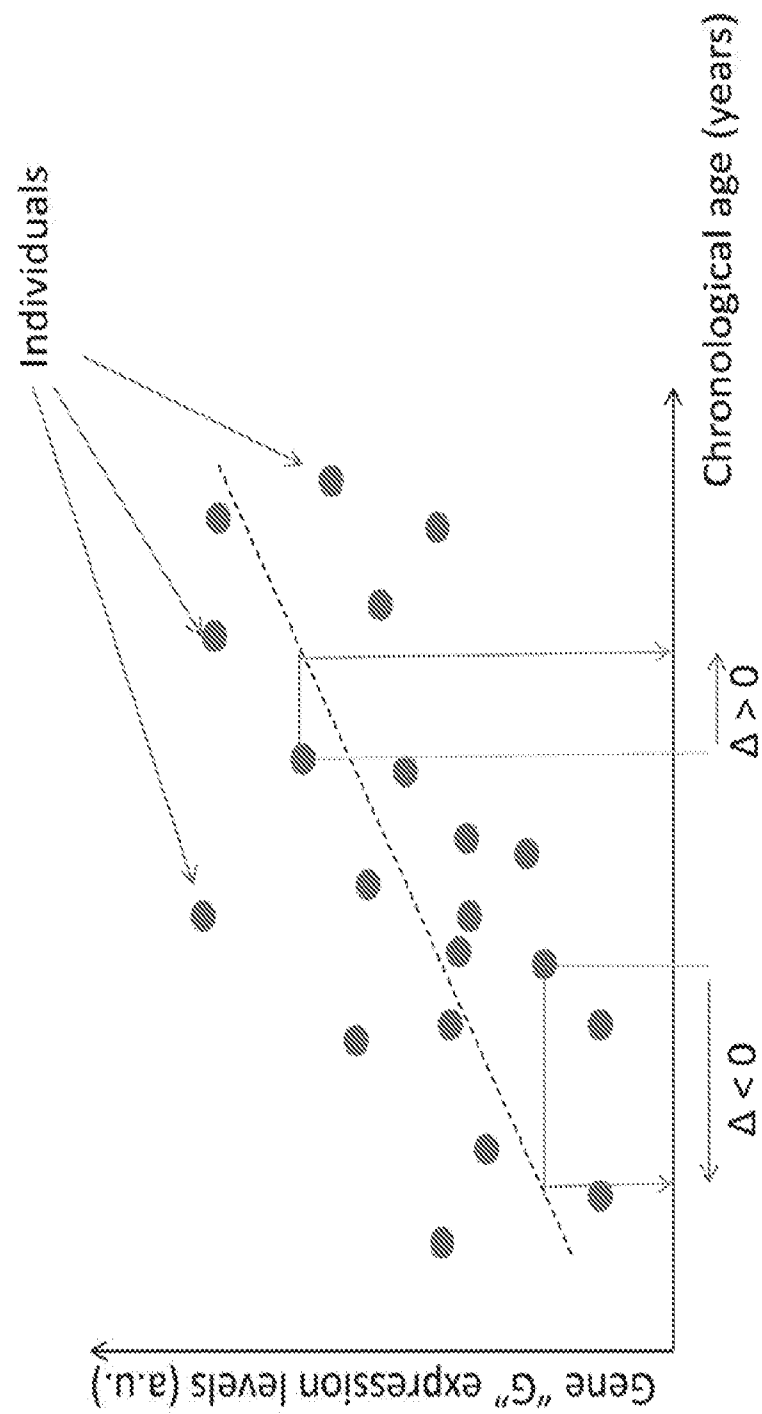


Fig. 19

**Fig. 20**

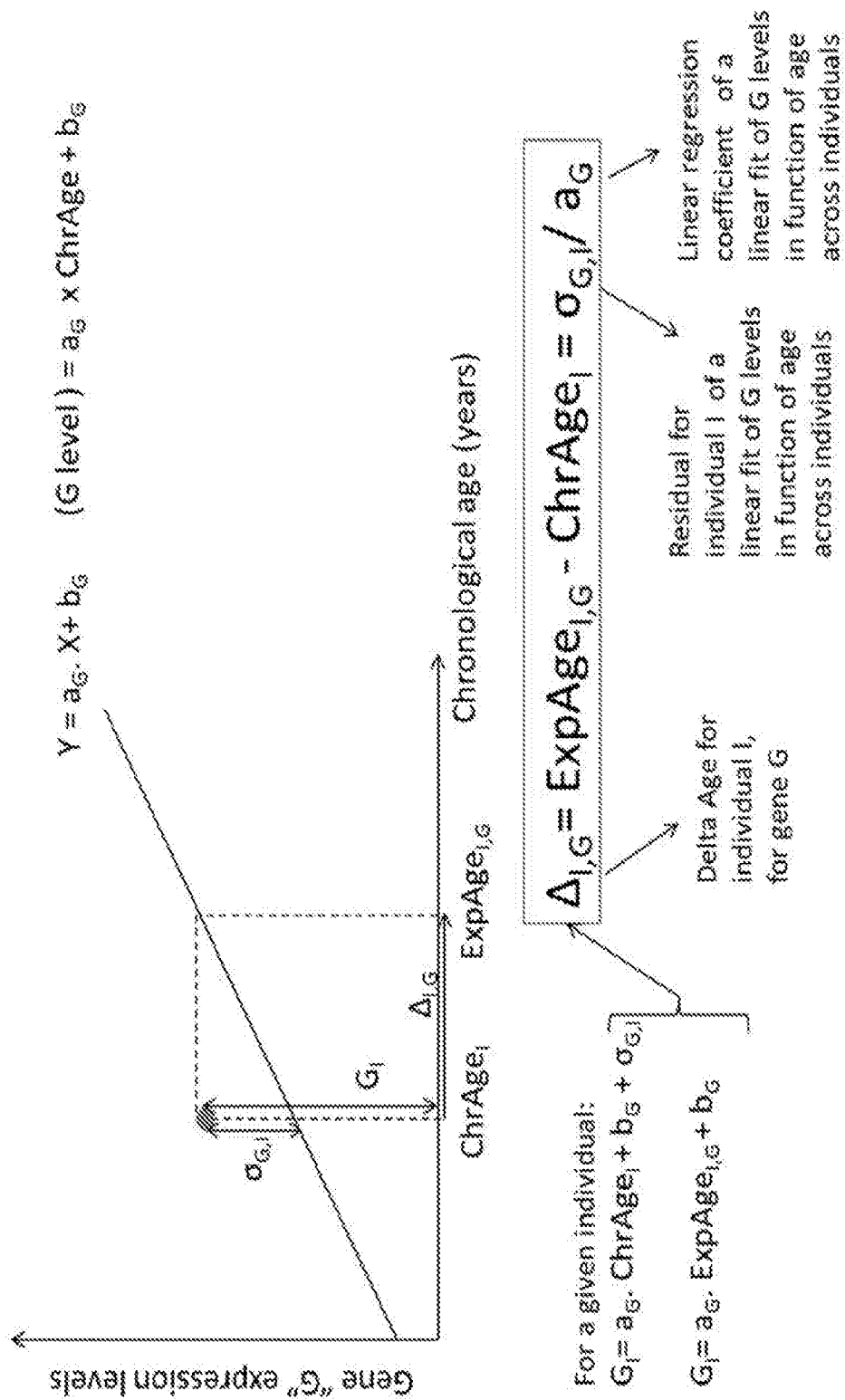


Fig. 21

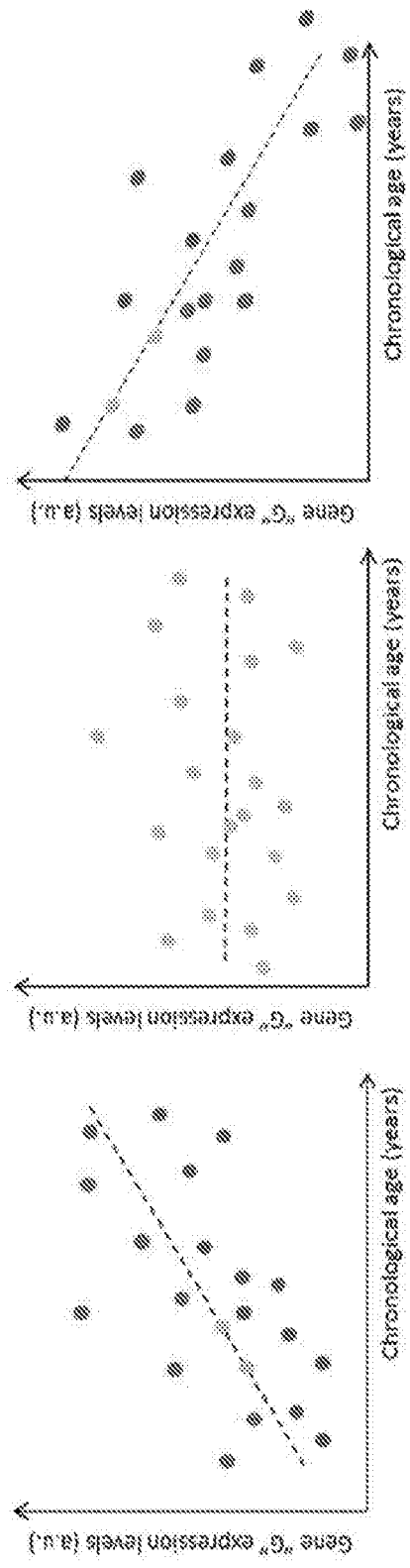


Fig. 22

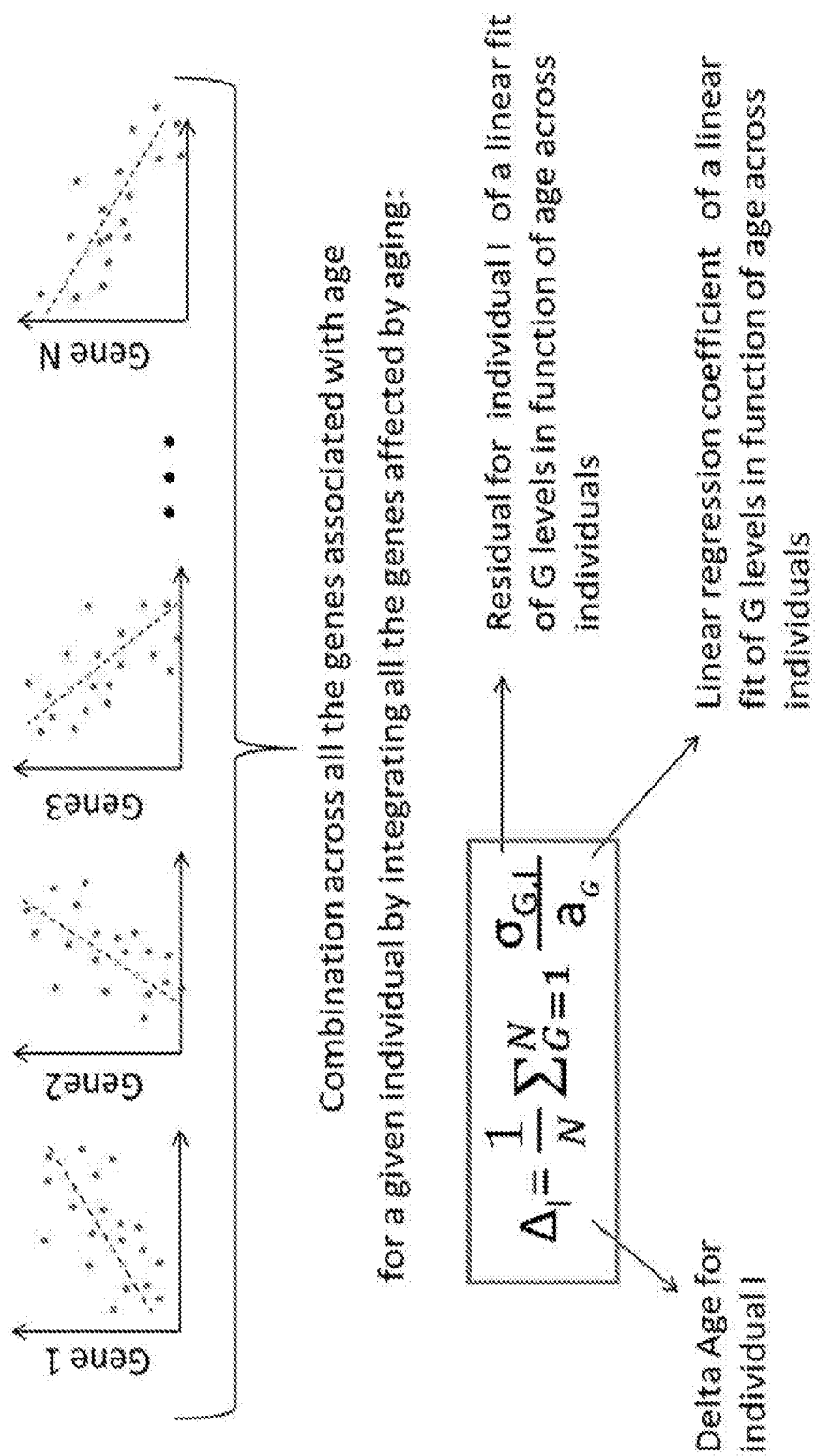


Fig. 23

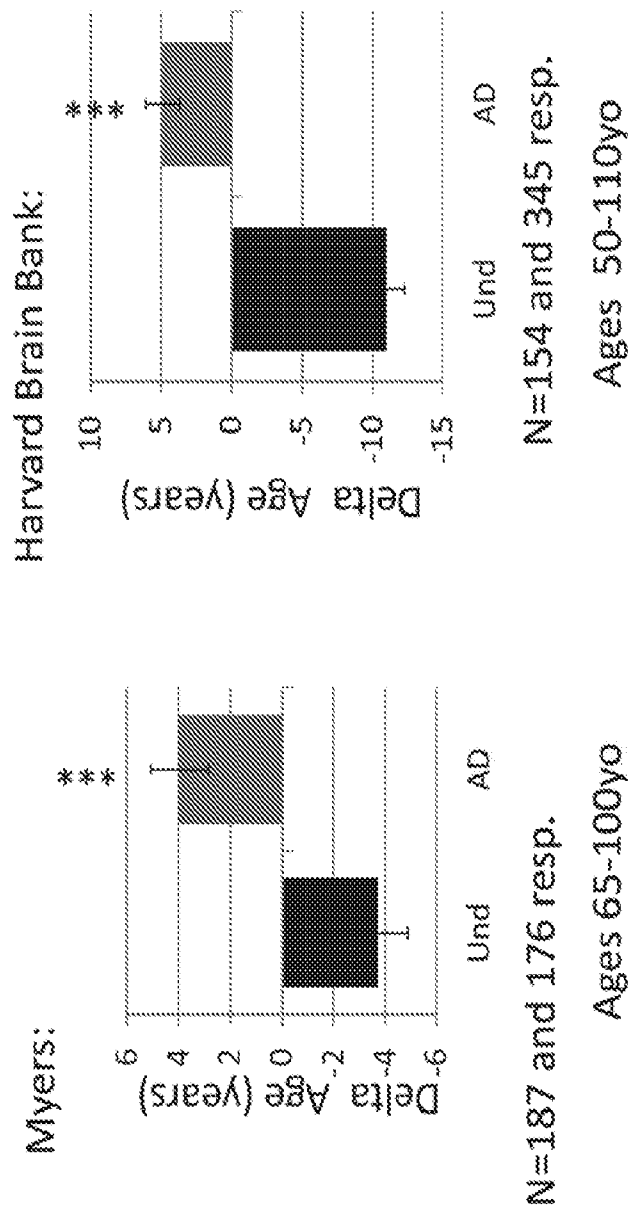


Fig. 24

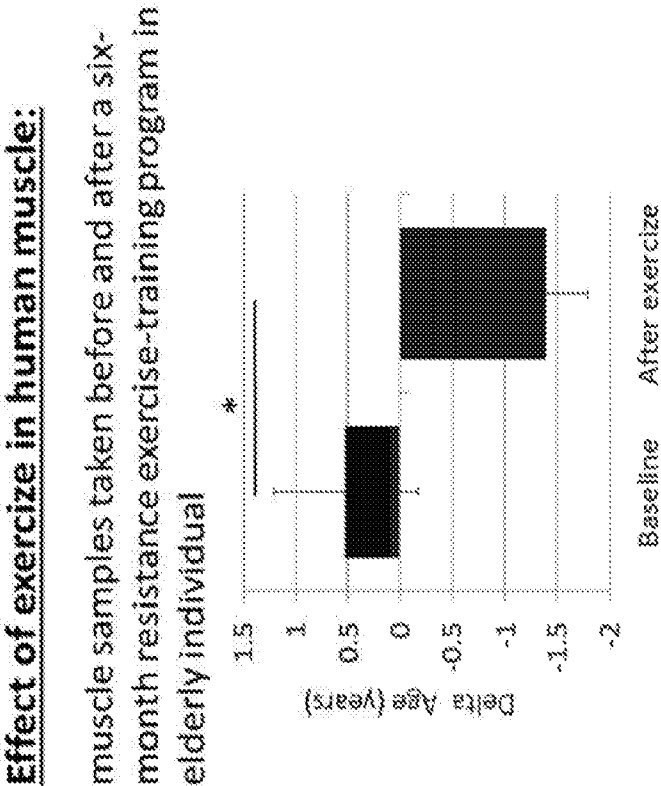


Fig. 25

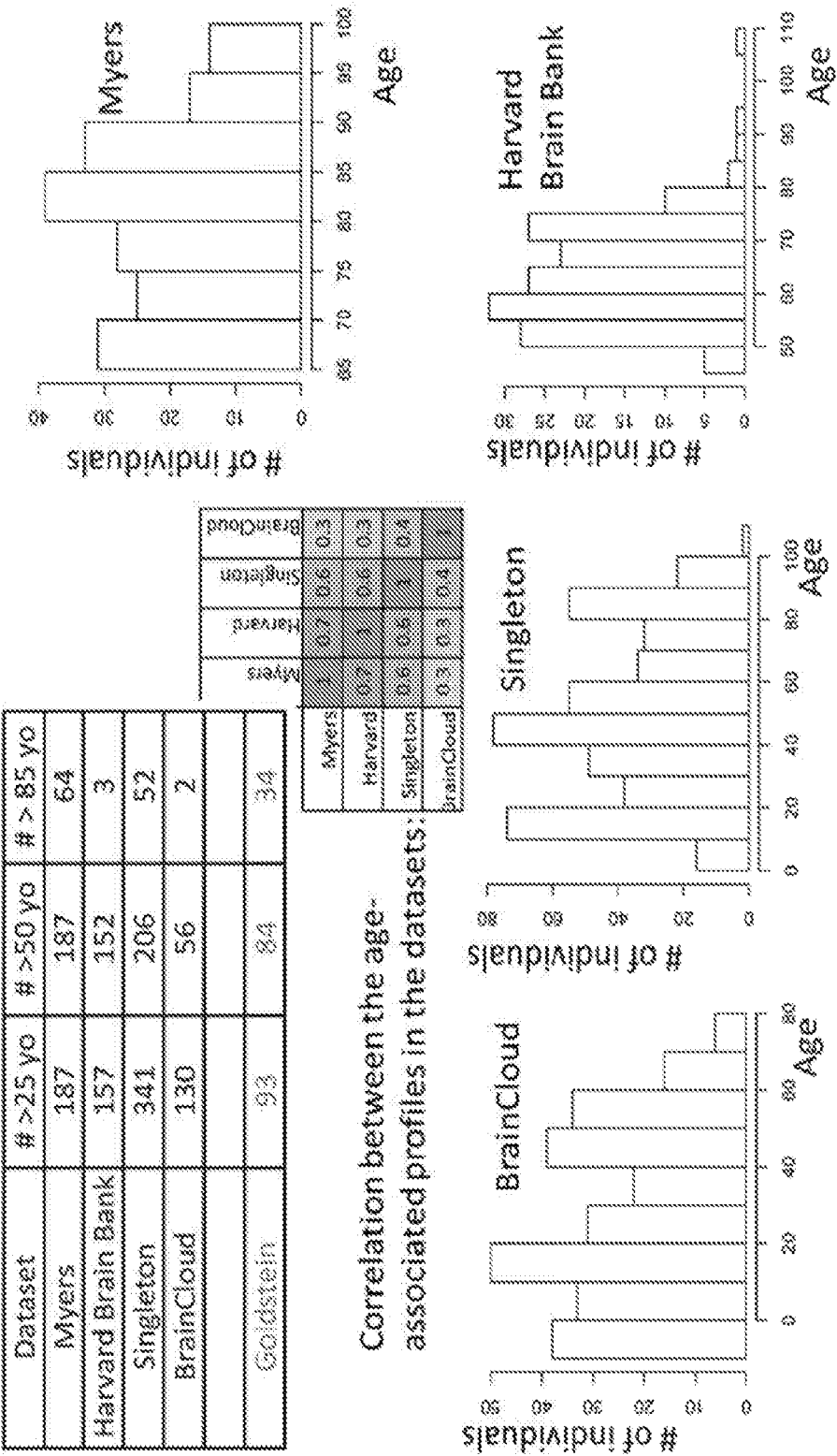


Fig. 26

ABCA1	ATP binding cassette subfamily A member 1(ABCA1)
ADCYAP1R1	ADCYAP receptor type I(ADCYAP1R1)
C4A	complement C4A (Rodgers blood group)(C4A)
CCL2	C-C motif chemokine ligand 2(CCL2)
CCL3	C-C motif chemokine ligand 3(CCL3)
CCL4	C-C motif chemokine ligand 4(CCL4)
CCL5	C-C motif chemokine ligand 5(CCL5)
CD14	CD14 molecule(CD14)
CD86	CD86 molecule(CD86)
CHL1	cell adhesion molecule L1 like(CHL1)
CLU	clusterin(CLU)
CNN3	calponin 3(CNN3)
FCGR1A	Fc fragment of IgG receptor Ia(FCGR1A)
FCN2	ficolin 2(FCN2)
GBG11	globoside alpha-1,3-N-acetylgalactosaminyltransferase 1(GBG11)
HSPA1A	heat shock protein family A (Hsp70) member 1A(HSPA1A)
IGF2	insulin like growth factor 2(IGF2)
IL18	interleukin 18(IL18)
IL1A	interleukin 1 alpha(IL1A)
ITGA9	integrin subunit alpha 9(ITGA9)
KCND1	potassium voltage-gated channel subfamily D member 1(KCND1)
KCTD12	potassium channel tetramerization domain containing 12(KCTD12)
KL	klotho(KL)
LACC1	laccase domain containing 1(LACC1)
LY96	lymphocyte antigen 96(LY96)
NR3C1	nuclear receptor subfamily 3 group C member 1(NR3C1)
PLPP1	phospholipid phosphatase 1(PLPP1)
PPP1R15A	protein phosphatase 1 regulatory subunit 15A(PPP1R15A)
PSD	pleckstrin and Sec7 domain containing(PSD)
PTPRG	protein tyrosine phosphatase, receptor type G(PTPRG)
PTPRM	protein tyrosine phosphatase, receptor type M(PTPRM)
SOC33	suppressor of cytokine signaling 3(SOC33)
SPTBN4	spectrin beta, non-erythrocytic 4(SPTBN4)
TGM1	transglutaminase 1(TGM1)
TLR1	toll like receptor 1(TLR1)
TLR6	toll like receptor 6(TLR6)
TMEM47	transmembrane protein 47(TMEM47)
TTR	transthyretin(TTR)
VAT1L	vesicle amine transport 1 like(VAT1L)

Fig. 27A

AK1	adenylate kinase 1(AK1)
ANG	angiogenin(ANG)
ARHGAP22	Rho GTPase activating protein 22(ARHGAP22)
BIN1	bridging integrator 1(BIN1)
C1QA	complement C1q A chain(C1QA)
CCND1	cyclin D1(CCND1)
CKB	creatine kinase B(CKB)
CTSL	cathepsin L(CTSL)
EGR2	early growth response 2(EGR2)
FGD2	FYVE, RhoGEF and PH domain containing 2(FGD2)
GAS6	growth arrest specific 6(GAS6)
GTF2H2	general transcription factor IIH subunit 2(GTF2H2)
HEXB	hexosaminidase subunit beta(HEXB)
IGF1	insulin like growth factor 1(IGF1)
KCNMA1	potassium calcium-activated channel subfamily M alpha 1(KCNMA1)
KHDRBS3	KH RNA binding domain containing, signal transduction associated 3(KHDRBS3)
LGMN	legumain(LGMN)
LTC4S	leukotriene C4 synthase(LTC4S)
MAFB	MAF bZIP transcription factor B(MAFB)
MYC	v-myc avian myelocytomatosis viral oncogene homolog(MYC)
NFIA	nuclear factor I A(NFIA)
NPL	N-acetylneuraminate pyruvate lyase(NPL)
PALD1	phosphatase domain containing, paladin 1(PALD1)
PDGFA	platelet derived growth factor subunit A(PDGFA)
PROS1	protein 5 (alpha)(PROS1)
PTGFRN	prostaglandin F2 receptor inhibitor(PTGFRN)
PTMS	parathymosin(PTMS)
RAB3IL1	RAB3A interacting protein like 1(RAB3IL1)
RASGRP3	RAS guanyl releasing protein 3(RASGRP3)
RTN4RL1	reticulon 4 receptor like 1(RTN4RL1)
SEMA6D	semaphorin 6D(SEMA6D)
SLC30A4	solute carrier family 30 member 4(SLC30A4)
STAB1	stabilin 1(STAB1)
TIMP2	TIMP metalloproteinase inhibitor 2(TIMP2)
TMEM64	transmembrane protein 64(TMEM64)
TREM2	triggering receptor expressed on myeloid cells 2(TREM2)
TRIM47	tripartite motif containing 47(TRIM47)
TSPAN7	tetraspanin 7(TSPAN7)
TTC28	tetratricopeptide repeat domain 28(TTC28)

Fig. 27B

SNP	Chr	Pos	Disease	Allele	Discovery+Replication			Disease		
					Zscore	P-value	N	Zscore	P-value	N
rs1990622	7	12283787	ALS/FTD	G	-9.63	5.75E-22	863	-6.27	3.74E-10	549
rs5848	17	42430244	GRN	T	5.50	3.89E-08	642	0.38	7.06E-01	555
rs8093731	18	29088958	AD	T	-2.48	1.30E-02	872	0.38	7.08E-01	555
rs34016896	3	160992864	PD	T	-2.36	1.85E-02	874	0.87	3.82E-01	557
rs10948363	6	47487762	AD	G	-7.23	2.55E-02	844	-1.52	1.28E-01	550
rs6656401	1	207692049	AD	A	2.15	3.16E-02	854	0.31	7.57E-01	557
rs7274581	20	55018260	AD	C	1.97	4.86E-02	855	-0.75	4.54E-01	320
rs3129882	6	32409530	PD	G	1.66	9.77E-02	819	-2.02	4.31E-02	554
rs11724635	4	15737101	PD	C	1.55	1.22E-01	828	-0.57	5.67E-01	554
rs3793947	11	83544472	PD	A	1.50	1.33E-01	722	1.01	3.11E-01	554
rs11158026	14	55348869	PD	T	-1.47	1.42E-01	841	-0.84	4.00E-01	557
rs704180	12	21994111	HS	G	-1.44	1.49E-01	858	0.19	8.52E-01	557
rs4698412	4	15737348	PD	G	1.44	1.50E-01	822	0.00	1.00E+00	555
rs4147929	19	1063443	AD	A	1.41	1.59E-01	641	1.62	1.06E-01	557
rs35749011	1	155135036	PD	A	1.40	1.61E-01	798	-0.26	7.99E-01	554
rs10792832	11	85867875	AD	A	-1.37	1.72E-01	862	-1.63	1.03E-01	549
rs10498633	14	92926952	AD	T	-1.36	1.75E-01	867	-1.11	2.67E-01	551
rs329648	11	133765367	PD	T	-1.28	2.02E-01	871	-0.27	7.87E-01	557
rs117896735	10	121536377	PD	A	1.26	2.06E-01	366	-0.60	5.50E-01	557
rs2823357	21	16914905	PD	A	-1.25	2.13E-01	833	-2.22	2.66E-02	557
rs199347	7	23293746	PD	G	-1.23	2.19E-01	855	0.29	7.72E-01	553
rs3899943	9	27543382	ALS/FTD	C	-1.22	2.23E-01	871	1.51	1.31E-01	548
rs3899942	9	27543281	ALS/FTD	T	-1.21	2.26E-01	872	1.01	3.11E-01	554
rs156429	7	23306020	PD	C	-1.15	2.51E-01	862	0.19	8.46E-01	554
rs190982	5	68223420	AD	G	1.13	2.58E-01	727	-0.35	7.23E-01	557
rs1491942	12	40620808	PD	G	-1.12	2.63E-01	361	-0.51	6.09E-01	543
rs2732703	17	44353222	AD	NA	-1.10	2.72E-01	314	-0.84	4.02E-01	540
rs34517613	17	26610252	ALS/FTD	T	1.07	2.84E-01	759	1.63	1.02E-01	557

Fig. 28

rs1980493	6	32363215	ALS/FTD	C	-1.05	2.94E-01	862	0.55	5.83E-01	553
rs12456492	18	40673380	PD	G	1.01	3.11E-01	867	0.32	7.52E-01	553
rs646776	1	109818530	ALS/FTD	C	0.99	3.22E-01	864	1.51	1.32E-01	546
rs115185635	3	87520857	PD	C	-0.94	3.45E-01	179	-1.31	1.89E-01	370
rs302668	11	87876911	ALS/FTD	C	0.92	3.59E-01	875	-2.57	1.02E-02	546
rs10838725	11	47557871	AD	C	0.85	3.93E-01	848	1.11	2.68E-01	557
rs9268856	6	32429719	ALS/FTD	A	0.79	4.33E-01	862	0.50	6.18E-01	552
rs1564282	4	852313	PD	T	0.78	4.36E-01	857	0.87	3.87E-01	554
rs34372695	1	156030037	PD	T	-0.77	4.39E-01	775	-1.05	2.93E-01	557
rs704178	12	21993478	HS	C	0.76	4.50E-01	862	1.49	1.37E-01	554
rs429358 (APOE E4)	19	45411941	AD	C	0.70	4.86E-01	746	-0.68	4.97E-01	554
rs2718058	7	37841534	AD	G	0.69	4.92E-01	819	0.16	8.72E-01	557
rs11060180	12	123303586	PD	G	0.63	5.31E-01	361	-0.36	7.16E-01	537
rs1476679	7	100004446	AD	C	-0.62	5.36E-01	866	-2.87	4.08E-03	552
rs1768208	3	39523003	ALS/FTD	T	0.60	5.48E-01	861	2.25	2.47E-02	555
rs9331896	8	27467686	AD	C	0.57	5.69E-01	799	0.81	4.18E-01	557
rs2102808	2	169117025	PD	T	0.57	5.71E-01	827	0.82	4.12E-01	557
rs11868035	17	17715101	PD	A	0.56	5.74E-01	789	0.20	8.40E-01	557
rs3747742	6	41162518	TREML2	C	0.55	5.80E-01	871	2.02	4.37E-02	557
rs35349669	2	234068476	AD	T	0.51	6.12E-01	714	-2.55	1.07E-02	557
rs356220	4	90641340	PD	T	-0.47	6.42E-01	365	1.14	2.54E-01	507
rs3865444	19	51727962	AD	A	-0.45	6.51E-01	831	0.22	8.30E-01	557
rs14235	16	31121793	PD	A	-0.42	6.77E-01	843	-0.25	8.01E-01	547
rs28834970	8	27195121	AD	C	-0.40	6.89E-01	857	0.20	8.40E-01	557
rs12817488	12	123296794	PD	G	0.34	7.32E-01	357	-0.74	4.58E-01	537
rs8070723	17	44081064	ALS/FTD	G	-0.32	7.52E-01	873	-1.51	1.32E-01	555
rs11771145	7	143110762	AD	A	-0.30	7.65E-01	557	0.97	3.35E-01	557
rs17125944	14	53400629	AD	C	-0.29	7.71E-01	872	-0.46	6.43E-01	554
rs10797576	1	232664611	PD	T	-0.27	7.84E-01	686	-0.79	4.29E-01	370
rs17649553	17	45994648	PD	T	-0.26	7.93E-01	871	-1.43	1.53E-01	555

Fig. 28 cont.

rs34311866	4	951947	PD	C	-0.25	8.06E-01	743	-0.57	5.71E-01	557
rs1474055	2	169110394	PD	T	0.24	8.07E-01	849	0.58	5.64E-01	557
rs11711441	3	182821275	PD	A	-0.24	8.07E-01	826	0.14	8.85E-01	557
rs983392	11	59923508	AD	G	-0.24	8.13E-01	852	0.71	4.77E-01	557
rs4889603	16	30982225	PD	G	0.23	8.19E-01	867	1.18	2.39E-01	513
rs356182	4	90626111	PD	G	-0.22	8.27E-01	788	-2.84	4.45E-03	541
rs12637471	3	182762437	PD	A	0.21	8.36E-01	875	0.69	4.90E-01	557
rs1876828	17	43911525	PD	T	-0.21	8.36E-01	870	-1.78	7.59E-02	557
rs11240572	1	205808013	PD	A	0.20	8.39E-01	841	-1.19	2.36E-01	544
rs6733839	2	127892810	AD	T	0.20	8.40E-01	472	-0.74	4.60E-01	552
rs9268877	6	32431147	ALS/FTD	A	0.19	8.53E-01	866	-0.48	6.29E-01	552
rs11218343	11	121435587	AD	C	0.17	8.65E-01	868	-1.11	2.66E-01	554
rs2414739	15	61994134	PD	G	0.13	8.93E-01	826	-0.84	4.03E-01	557
rs12608932	19	17752689	ALS/FTD	C	-0.11	9.13E-01	396	0.05	9.57E-01	556
rs7759295	6	41135850	TREM2	T	-0.10	9.18E-01	800	-0.56	5.77E-01	534
rs903603	9	27529316	ALS/FTD	A	-0.06	9.56E-01	841	-0.73	4.63E-01	553
rs9271192	6	32578530	AD	C	-0.03	9.76E-01	832	0.02	9.84E-01	557

Fig. 28 cont.

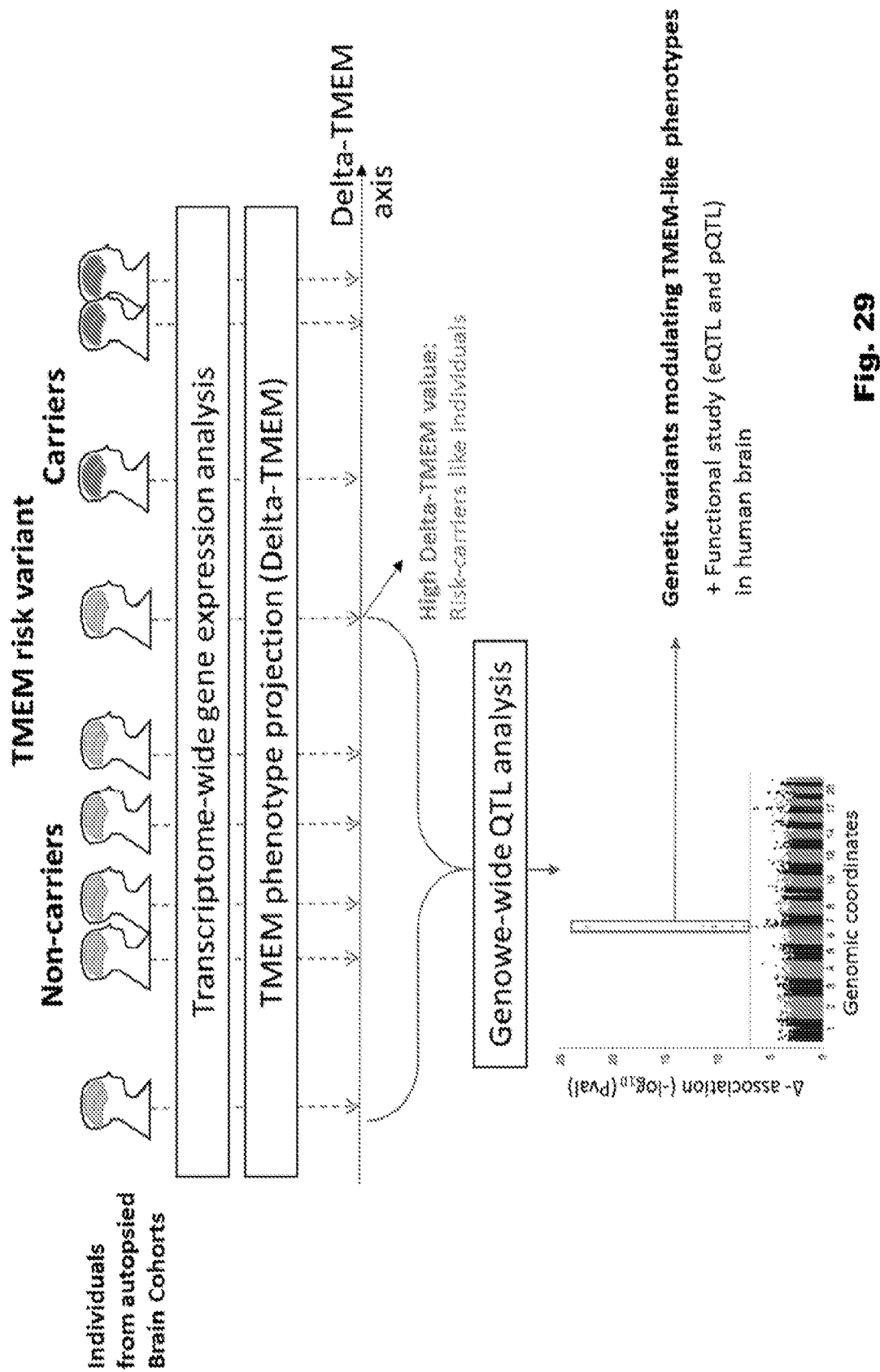


Fig. 29

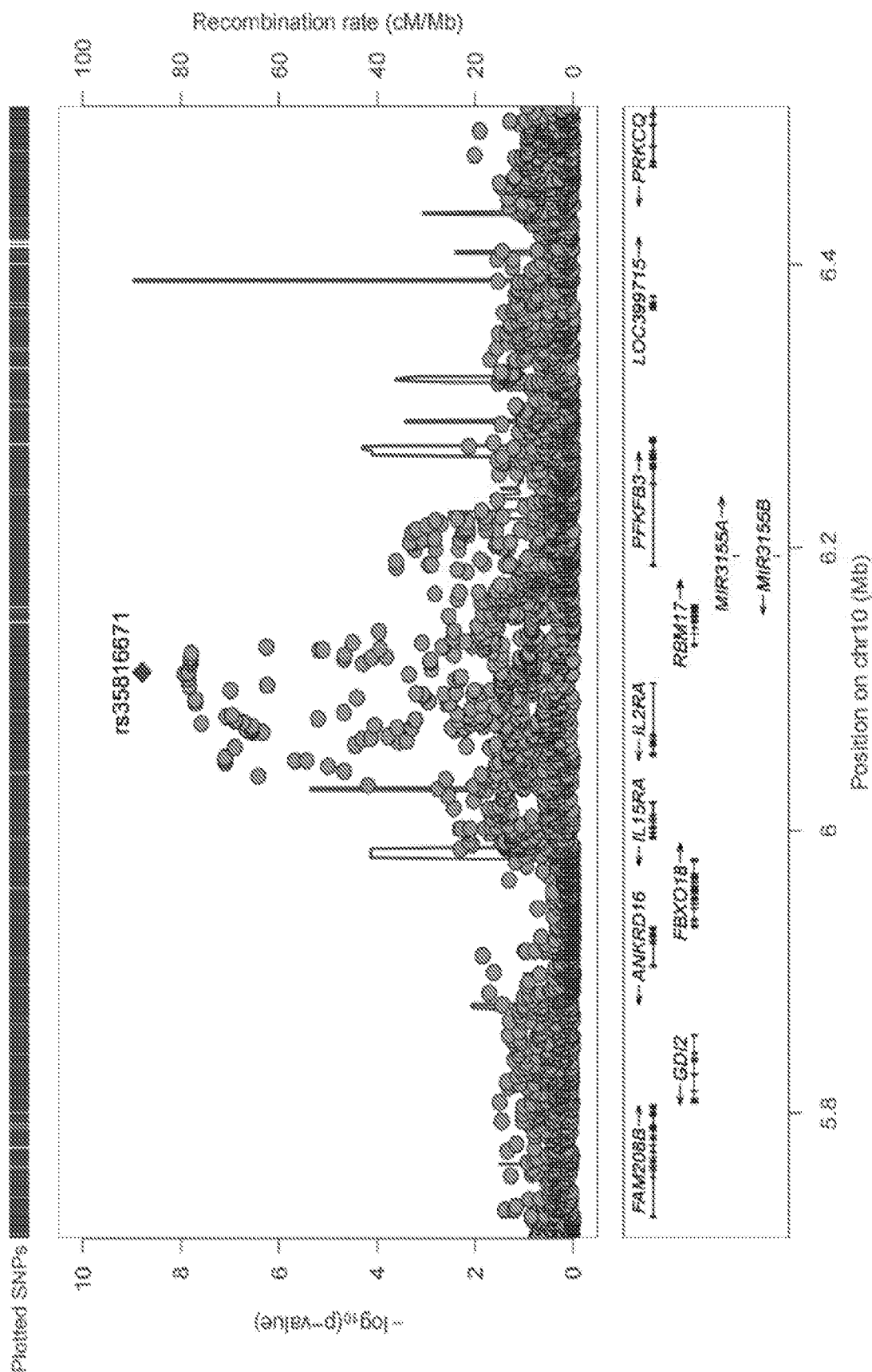


Fig. 30

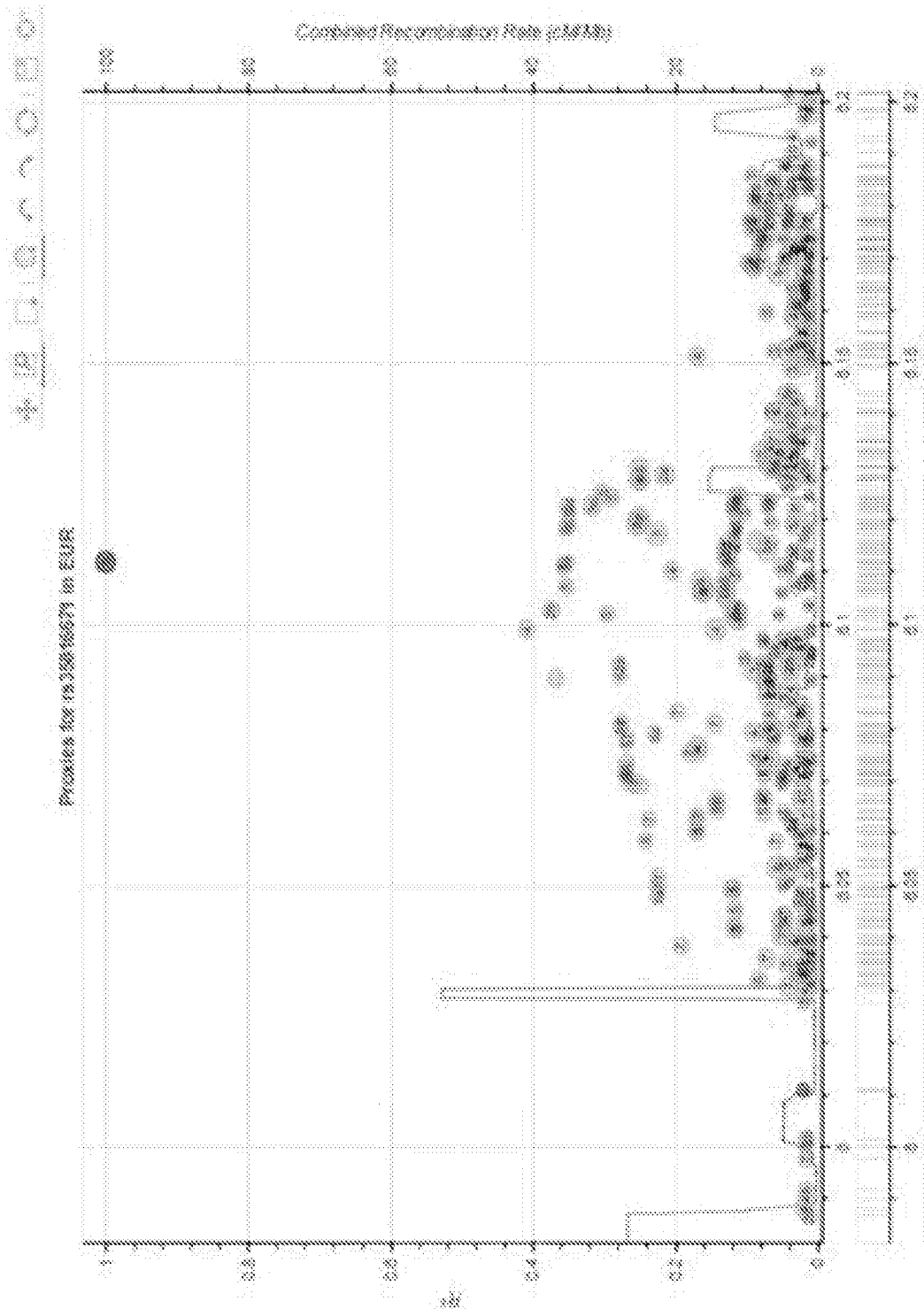


Fig. 31

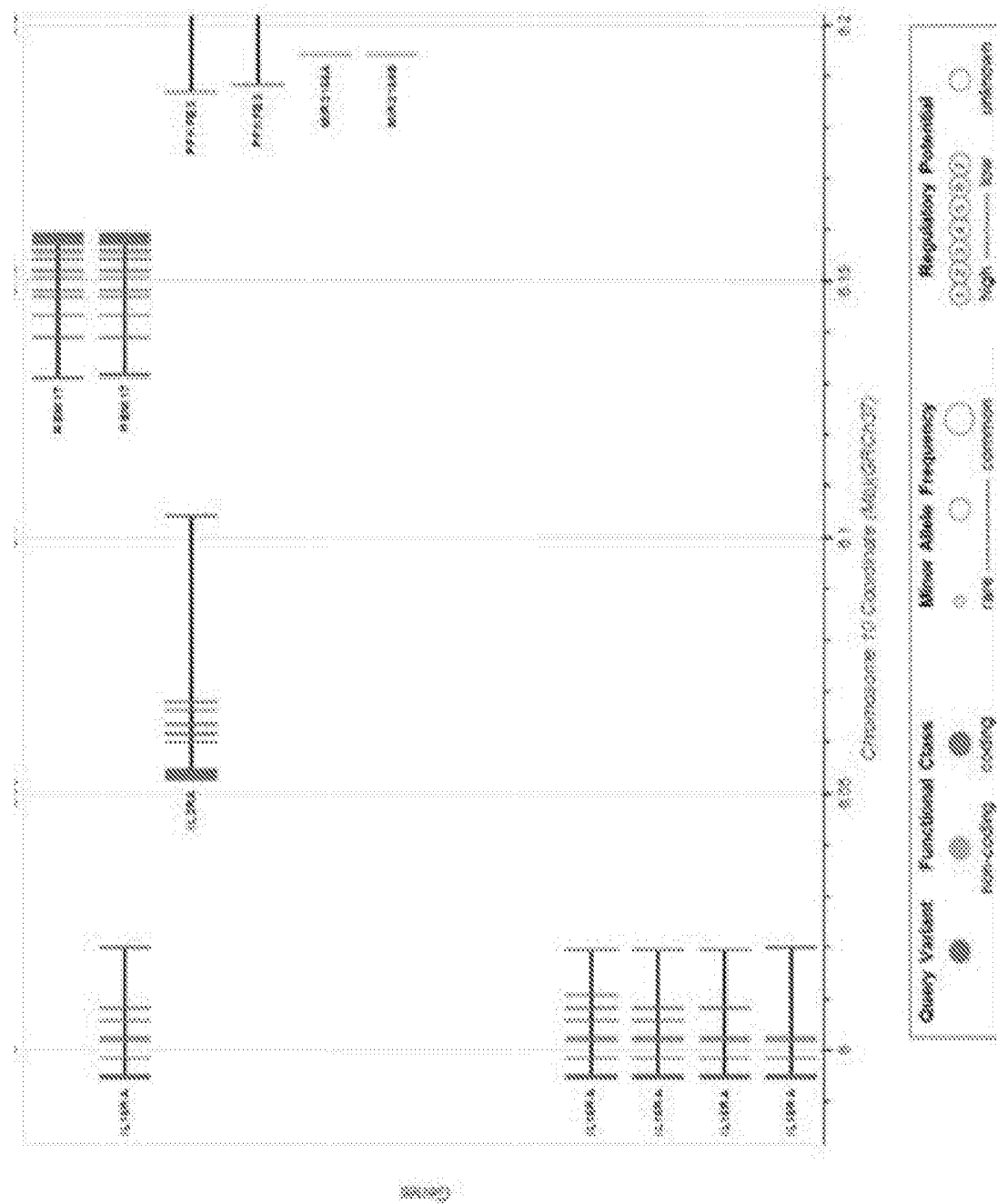


Fig. 31 cont.

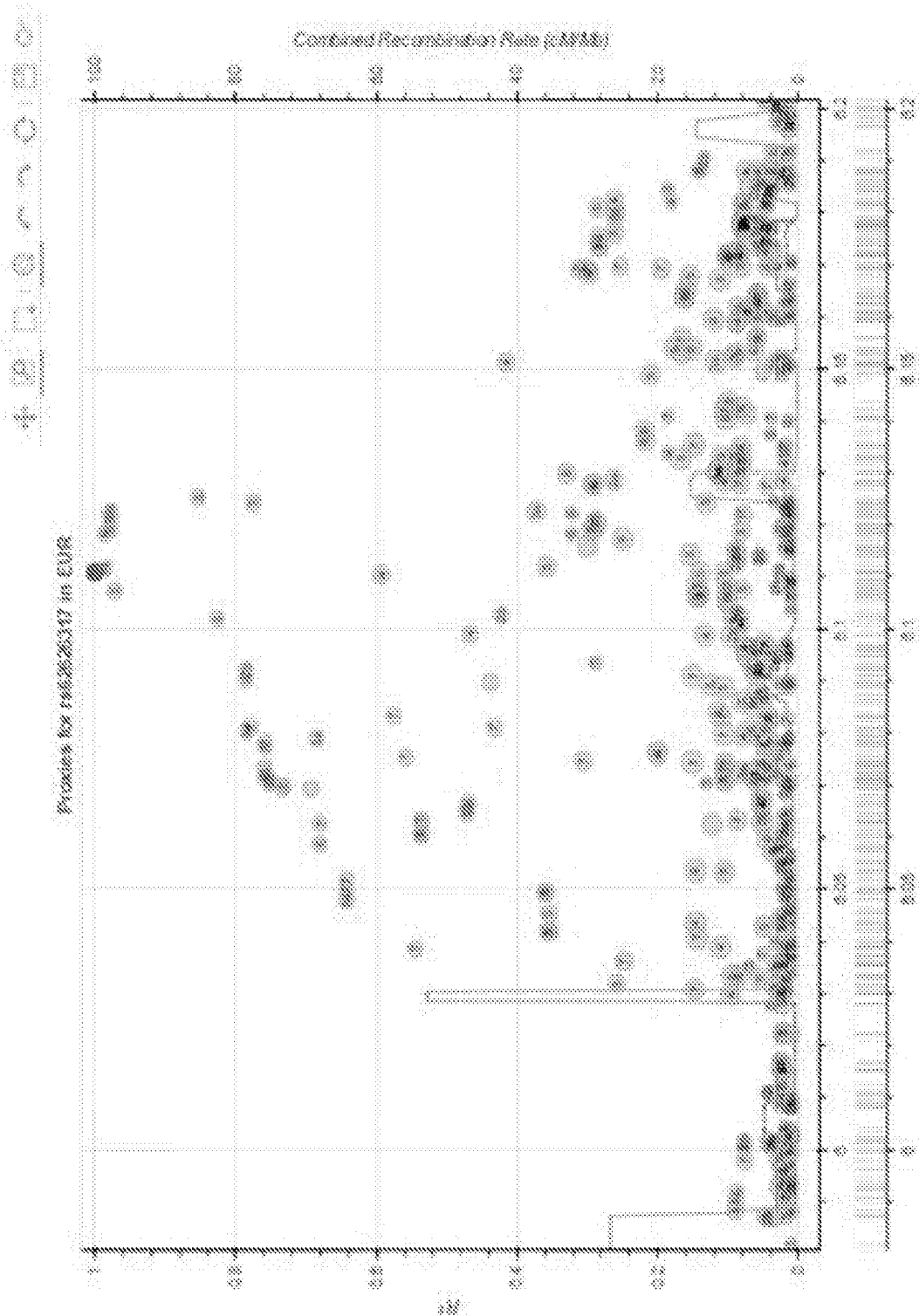


Fig. 31 cont.

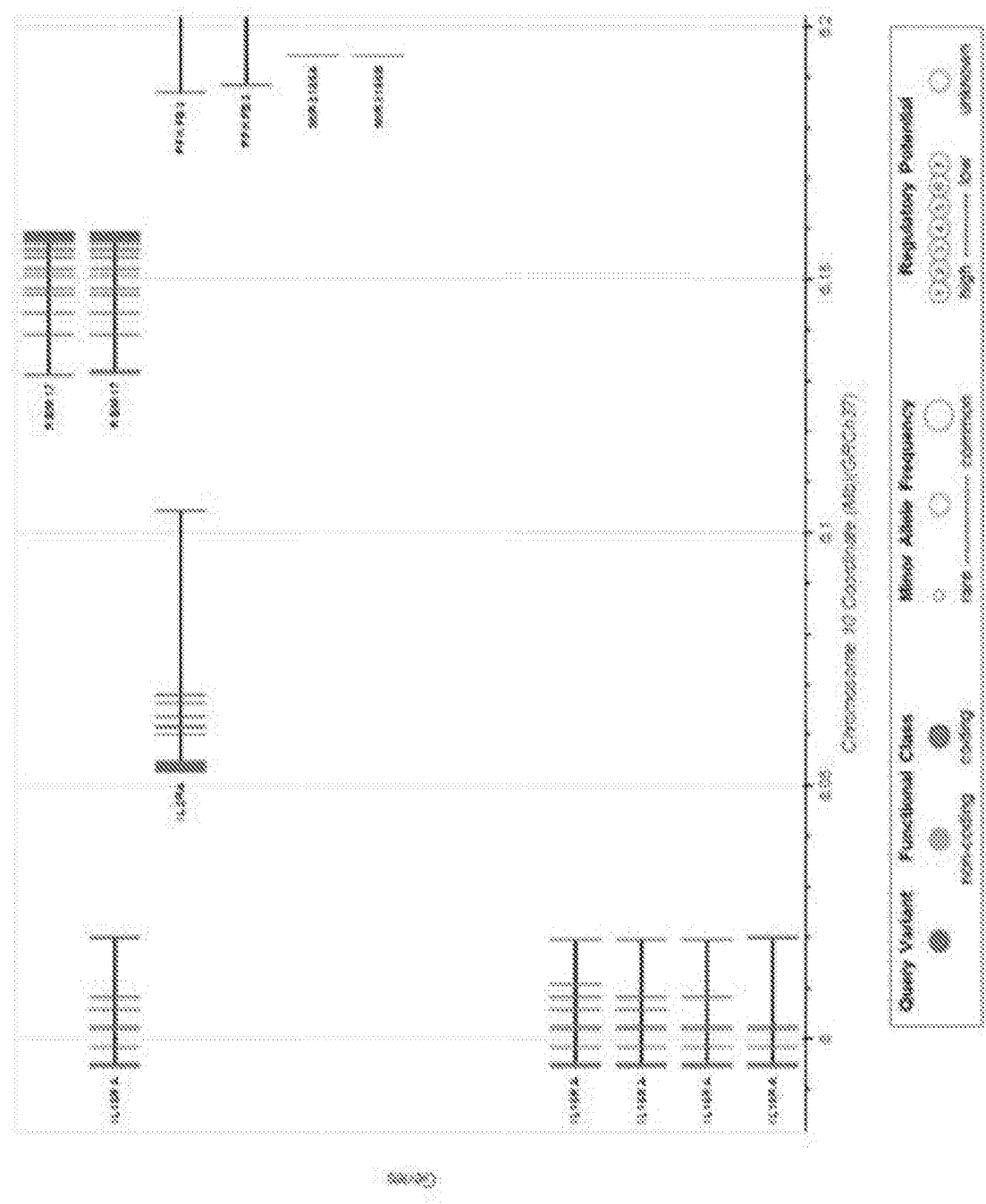


Fig. 31 cont.

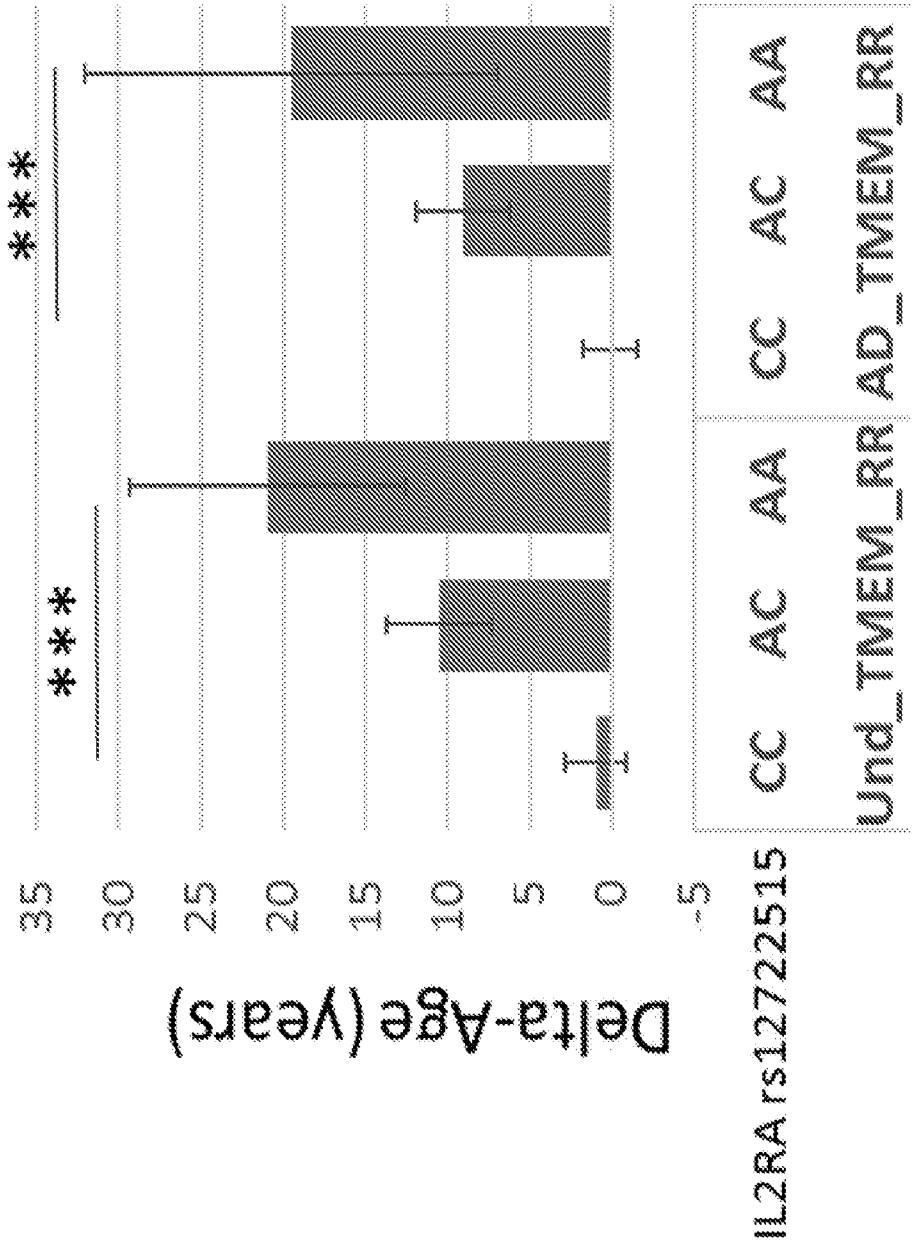


Fig. 32 N=150, 59, 8; 199, 61, 3

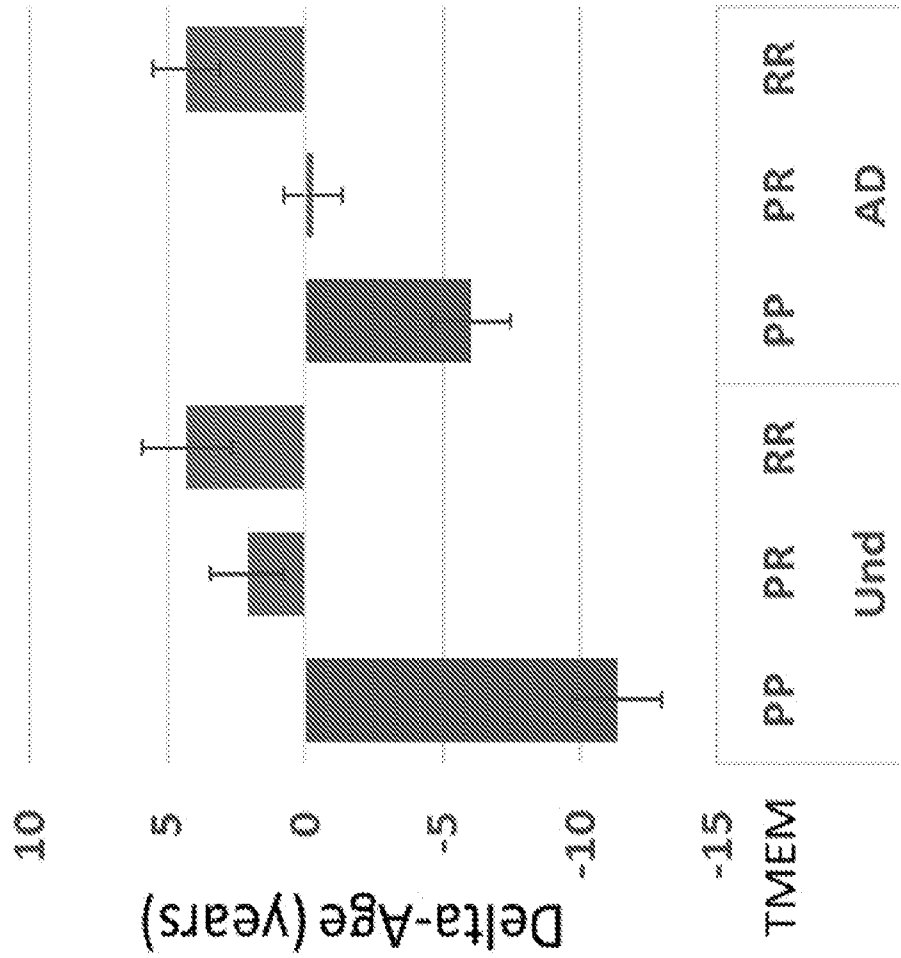


Fig. 33 n=147, 340, 217; 179, 381 and 263

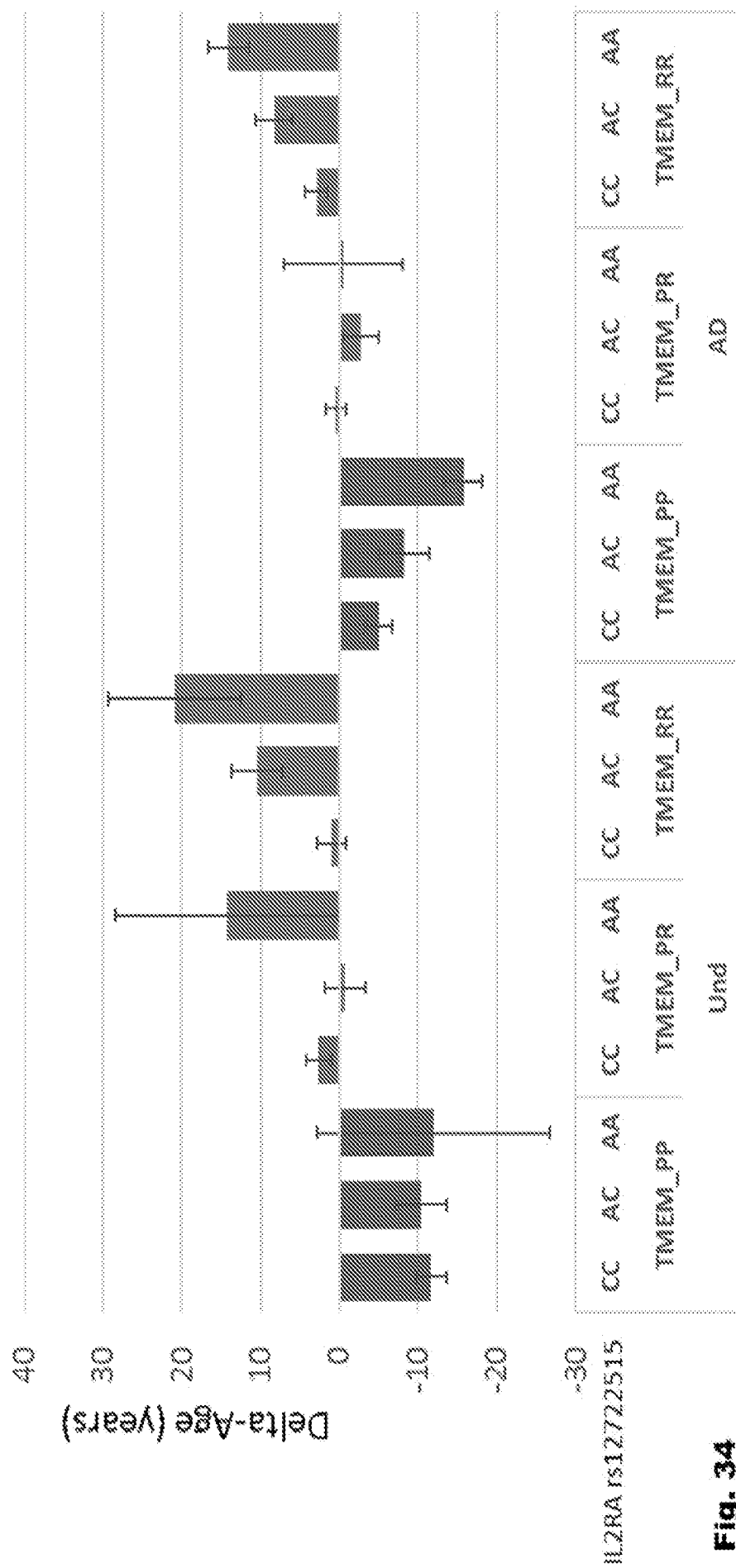


Fig. 34

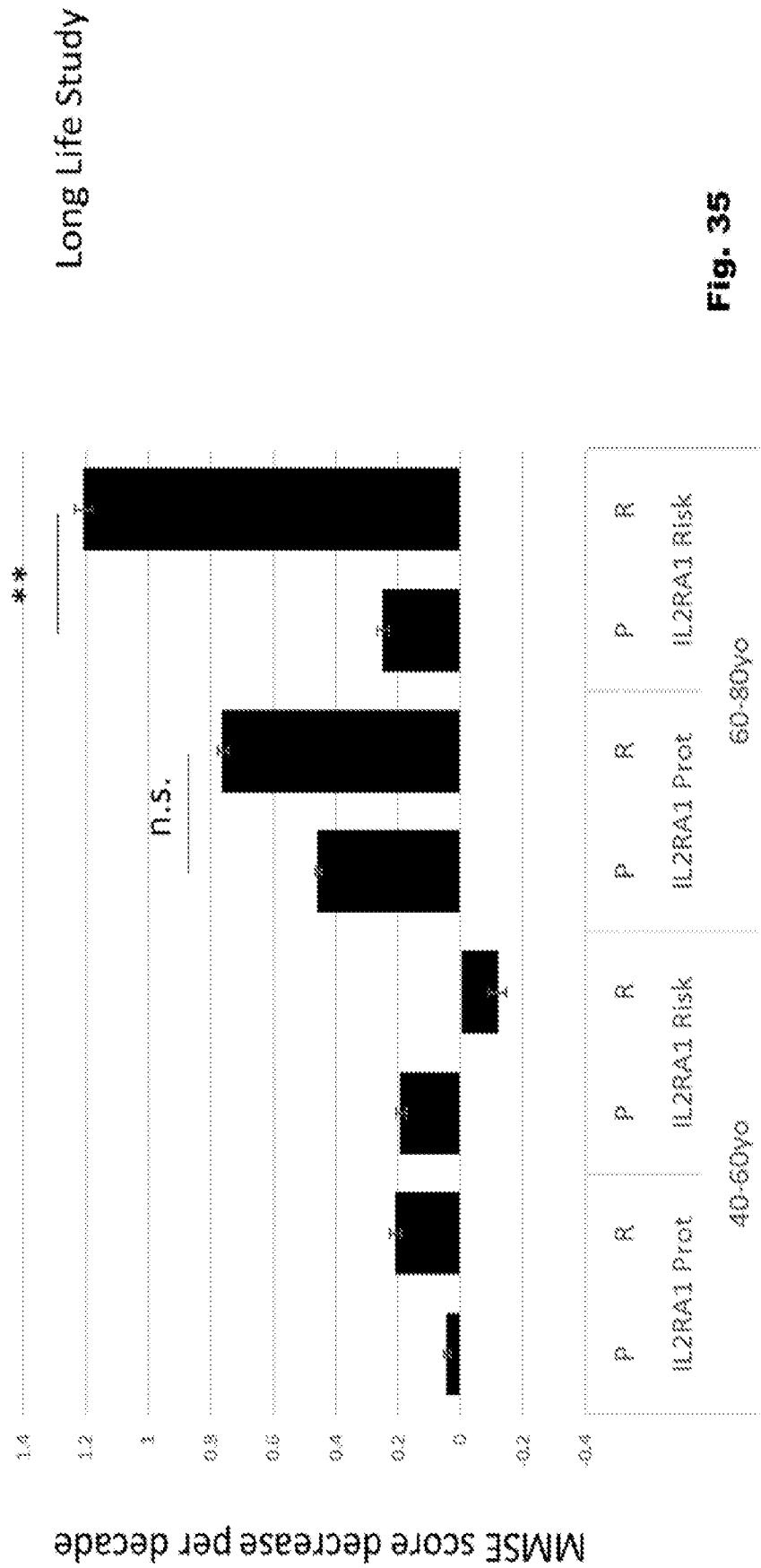


Fig. 35

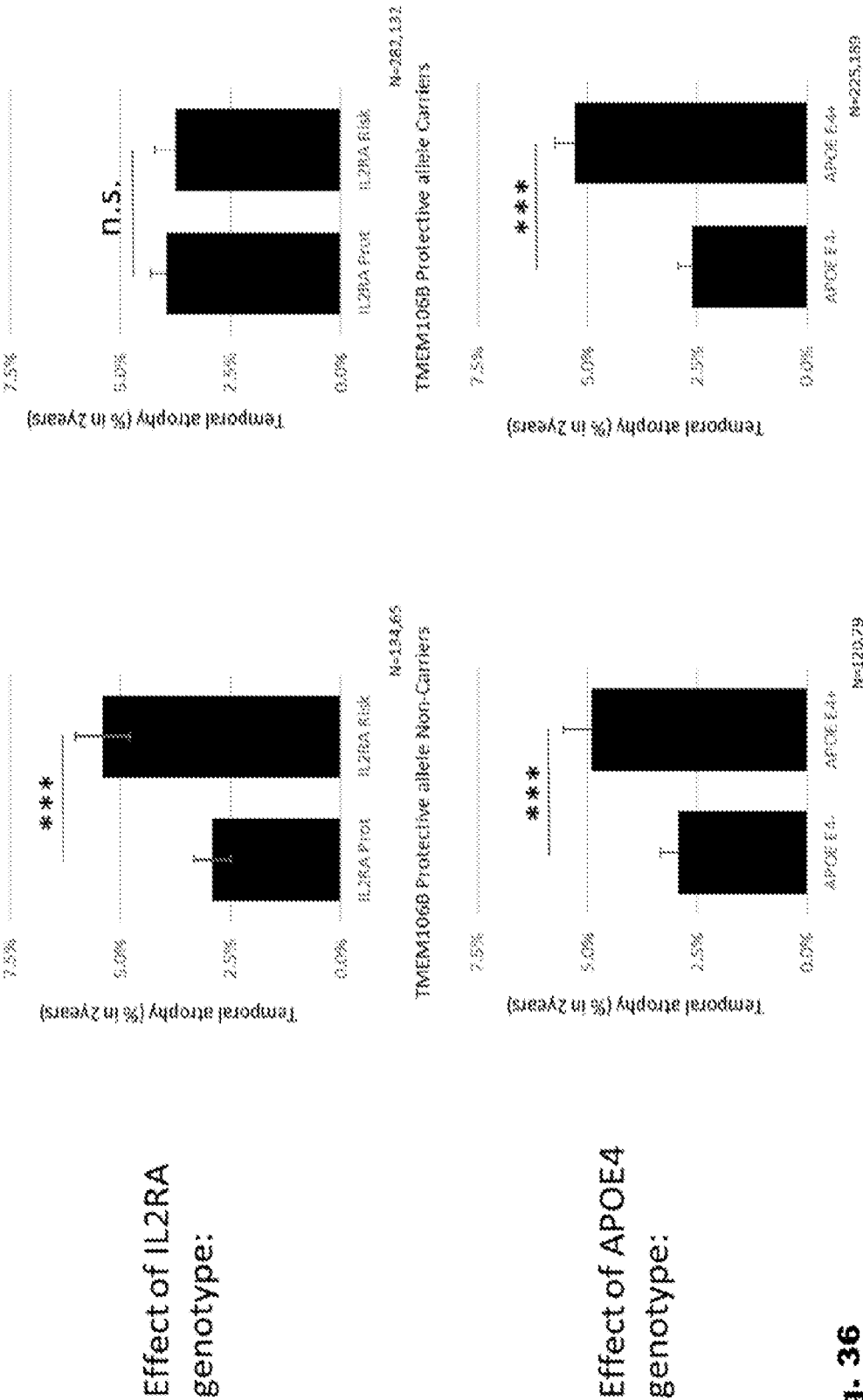
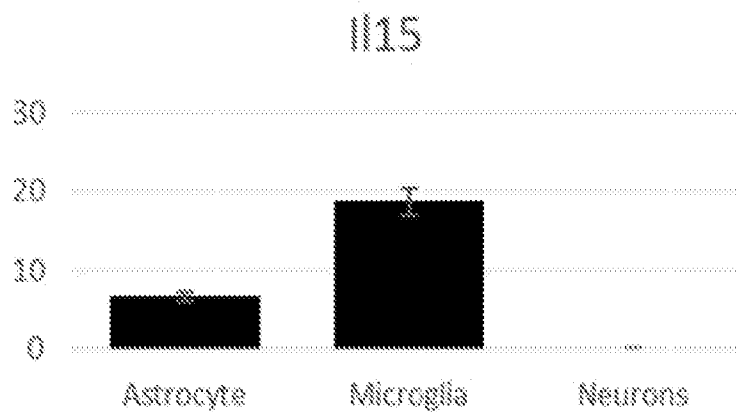
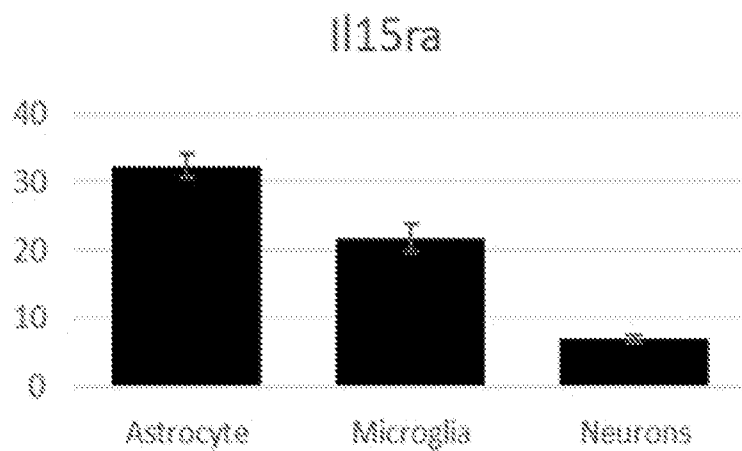


Fig. 36

	TMEM risk	Aging	AD
IL2	0.53	0.17	-0.05
IL2RA	-0.10	-0.33	-0.01
IL2RB	0.40	-3.45	-4.27
IL2RG	4.18	0.28	3.82
IL15	0.29	3.52	8.78
IL15RA	1.22	3.10	6.97

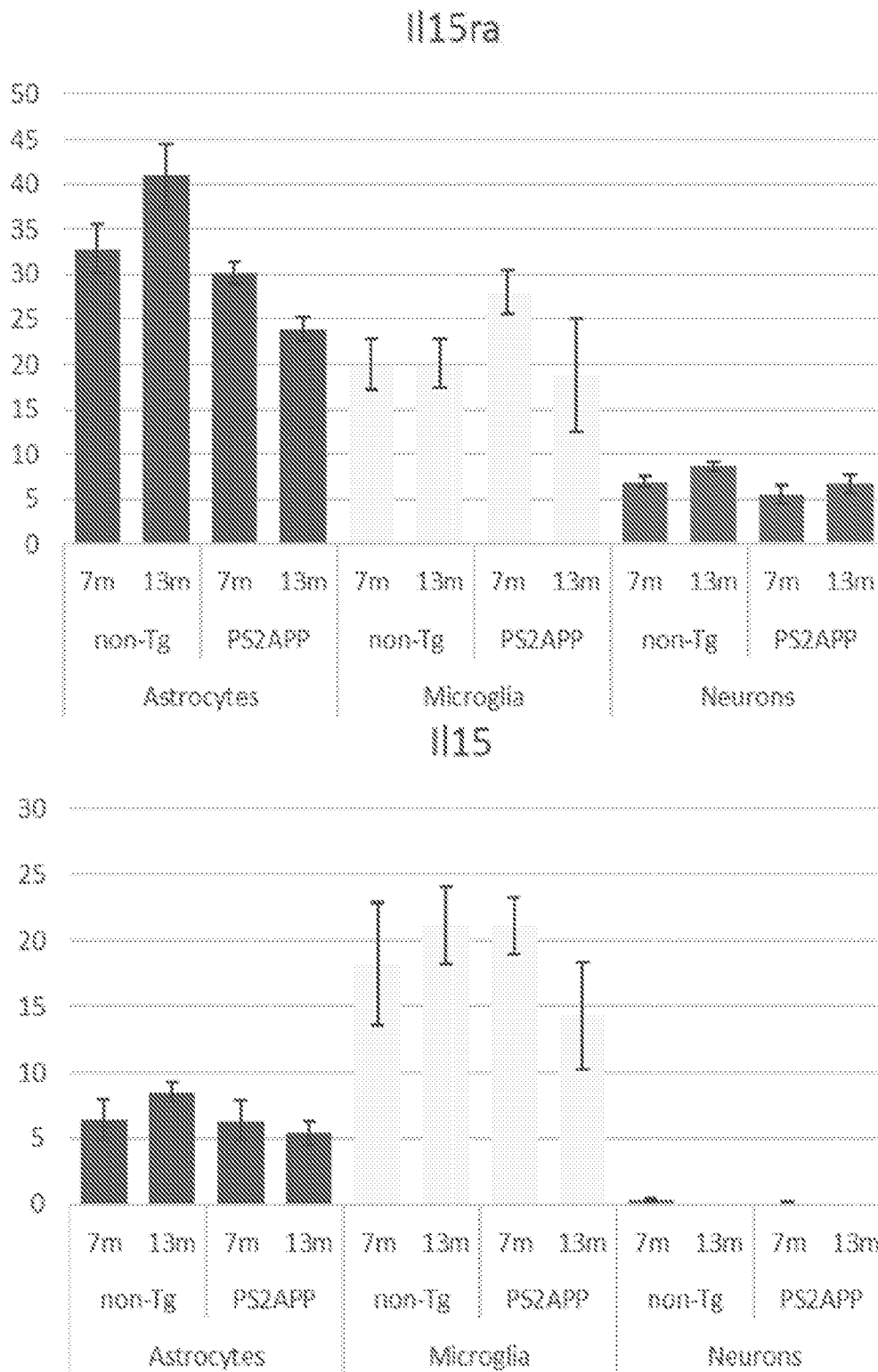
(Zscores)

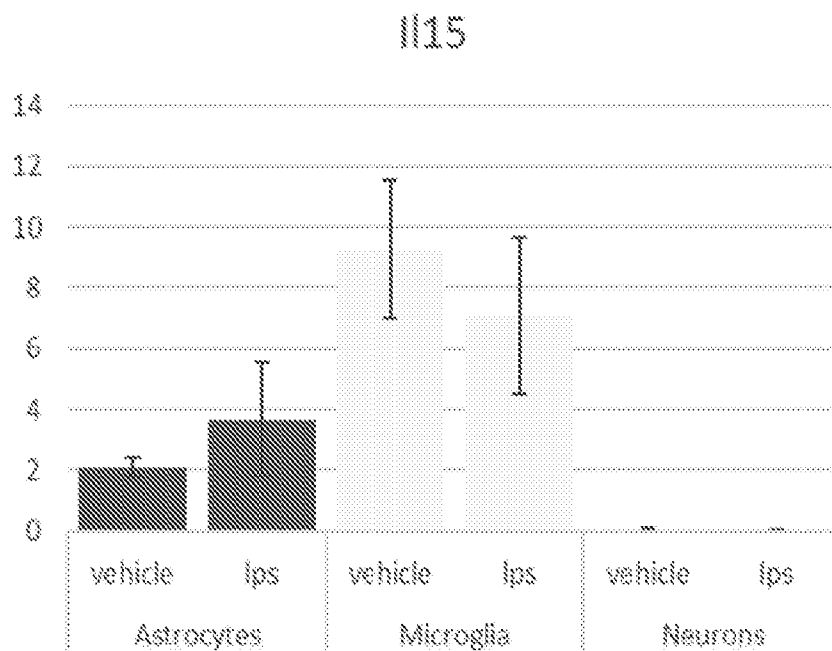
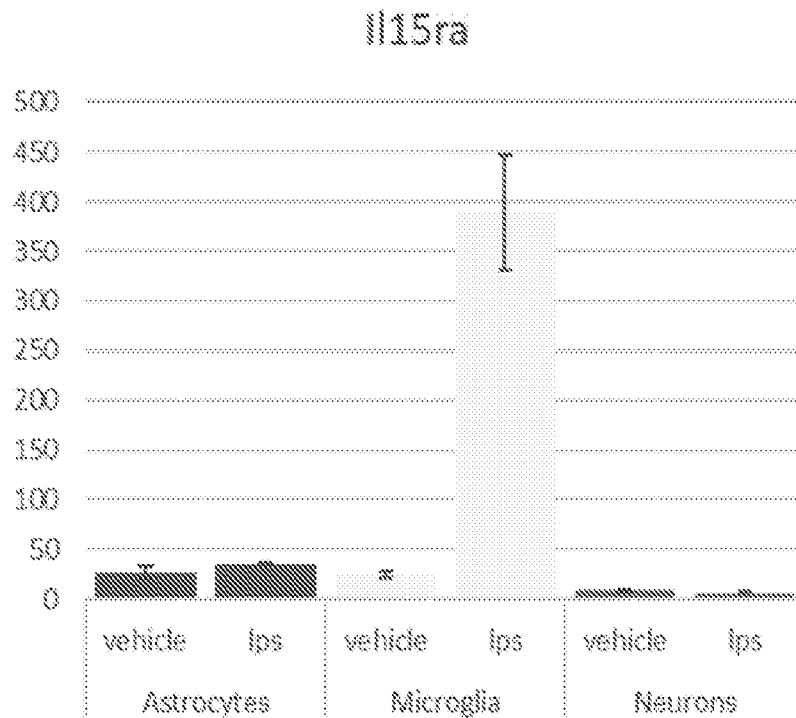
Fig. 37

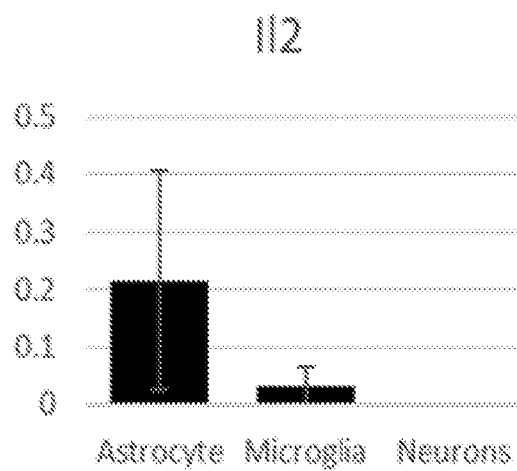
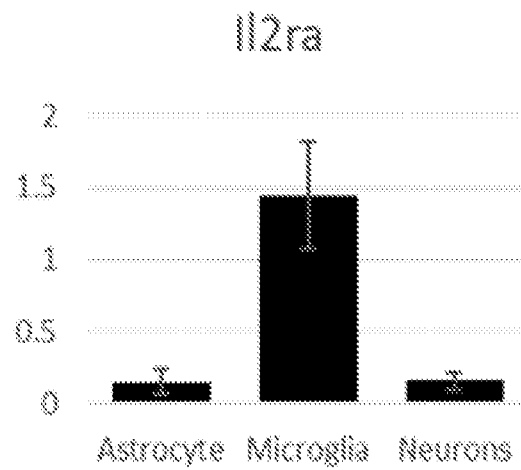


(Genentech, GSE75431)

Fig. 38

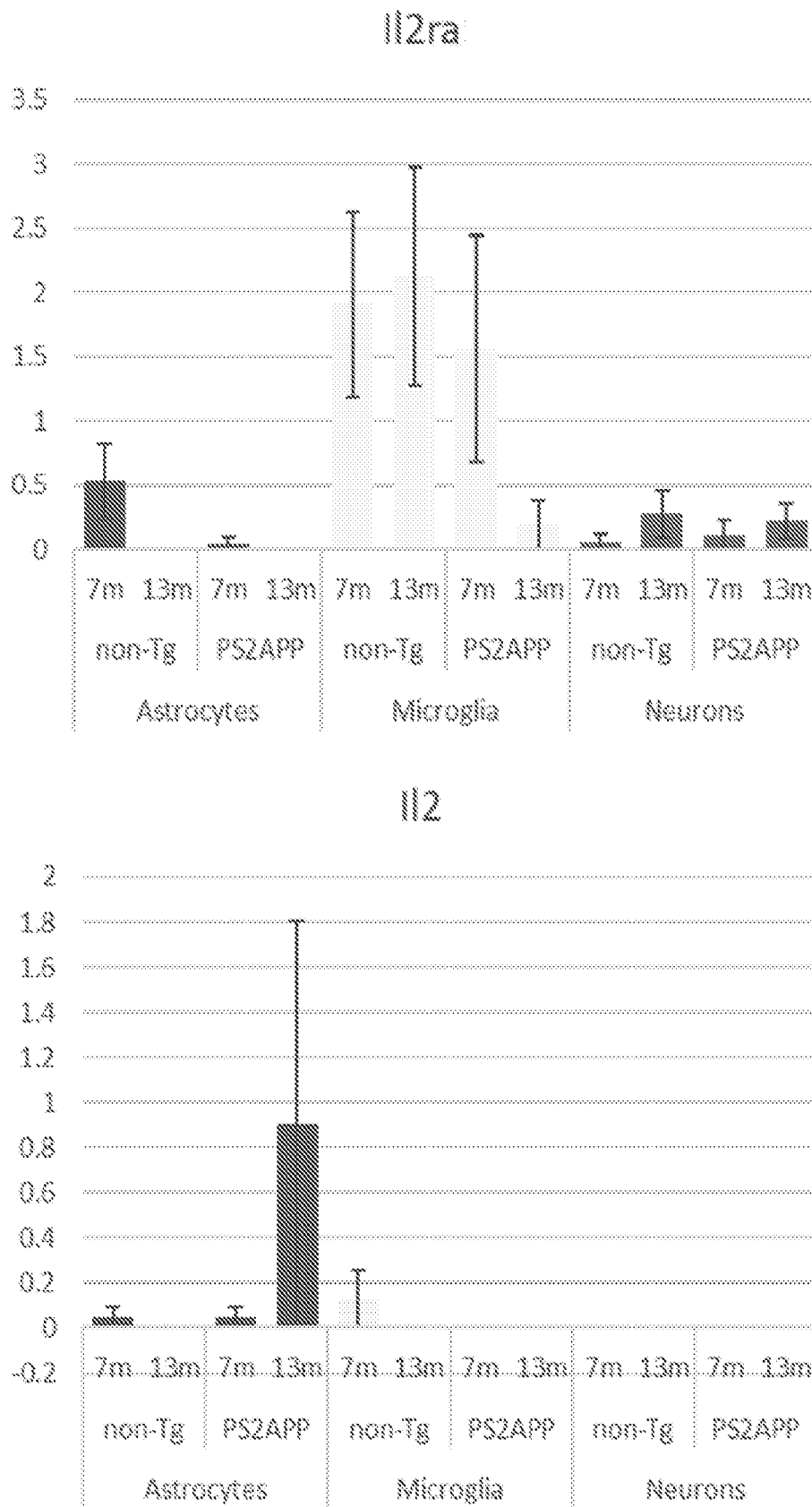
**Fig. 38 cont.**

**Fig. 38 cont.**



(Genentech, GSE75431)

Fig. 39

**Fig. 39 cont.**

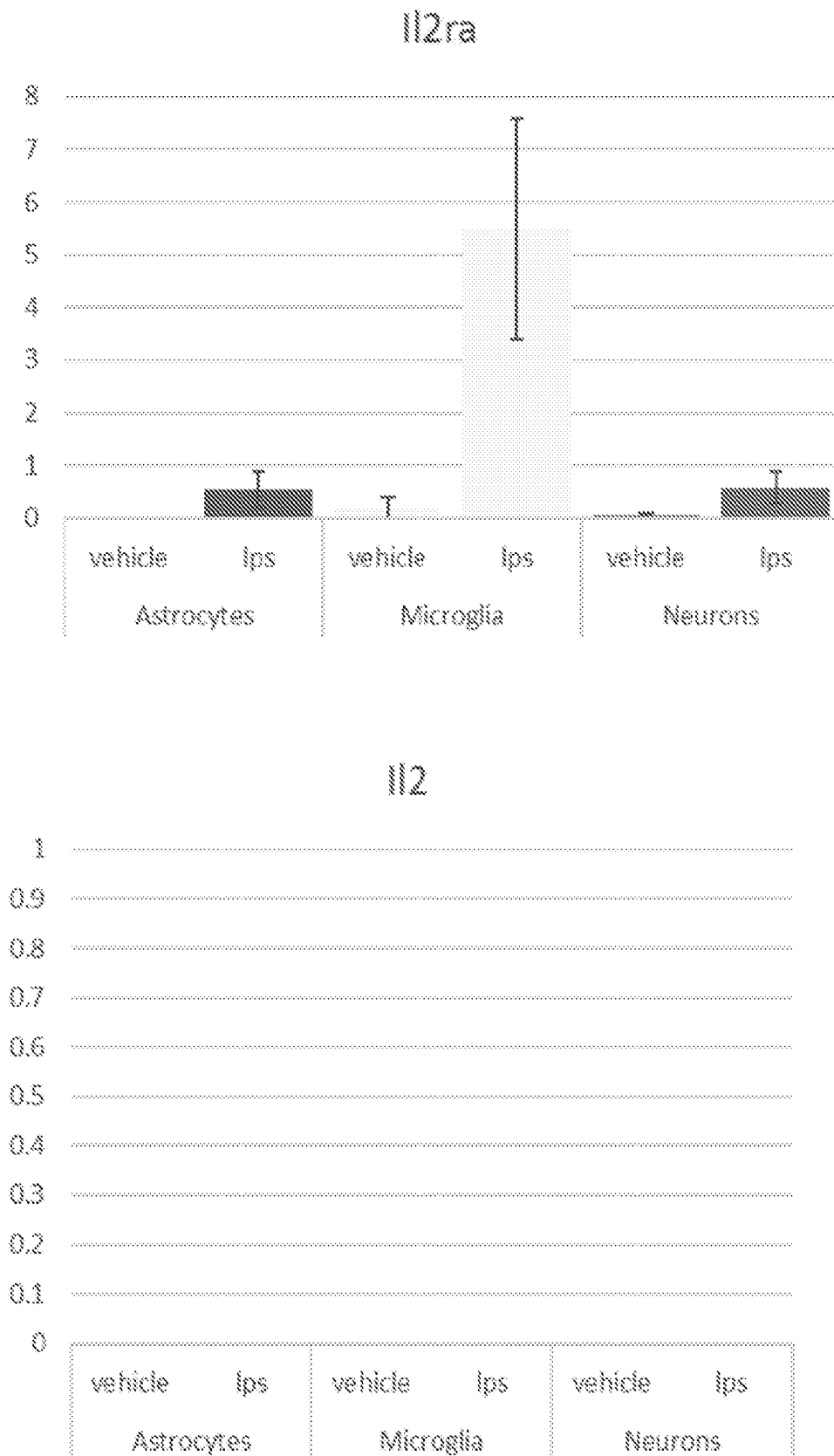
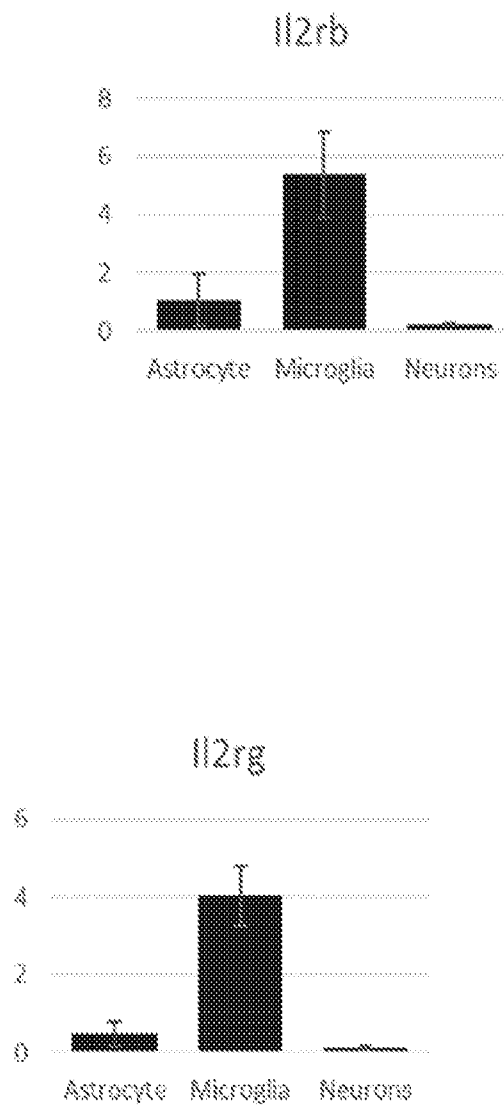
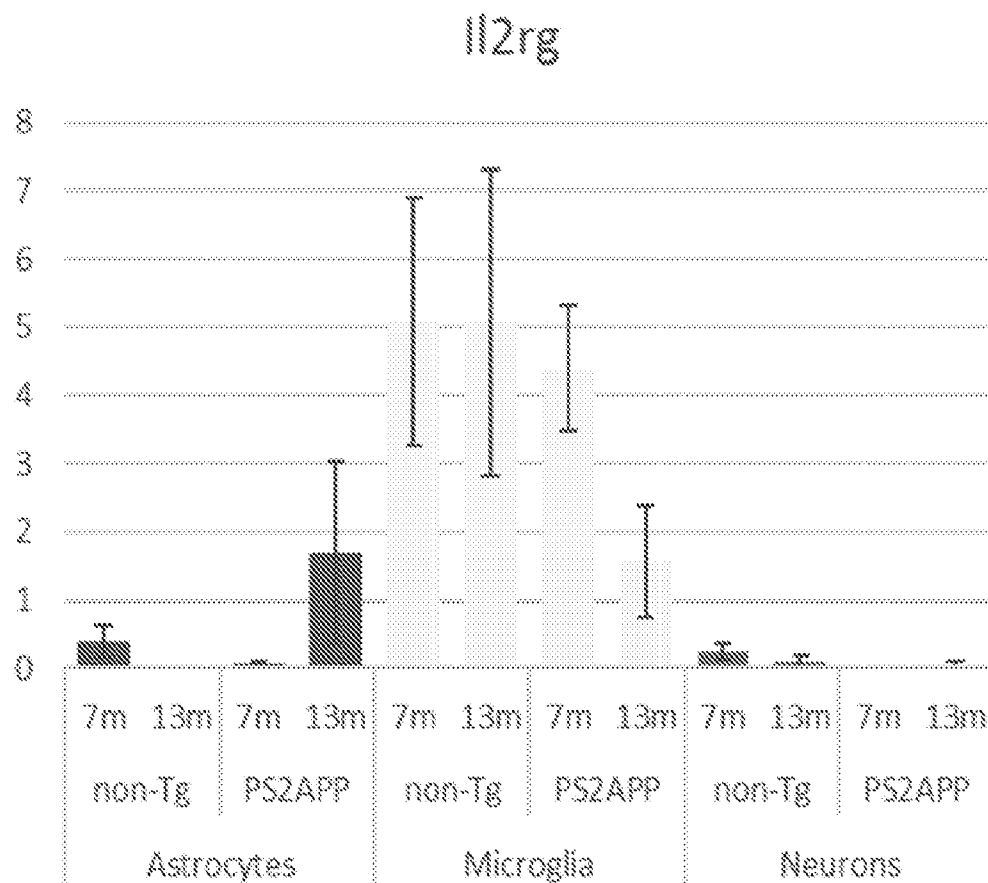
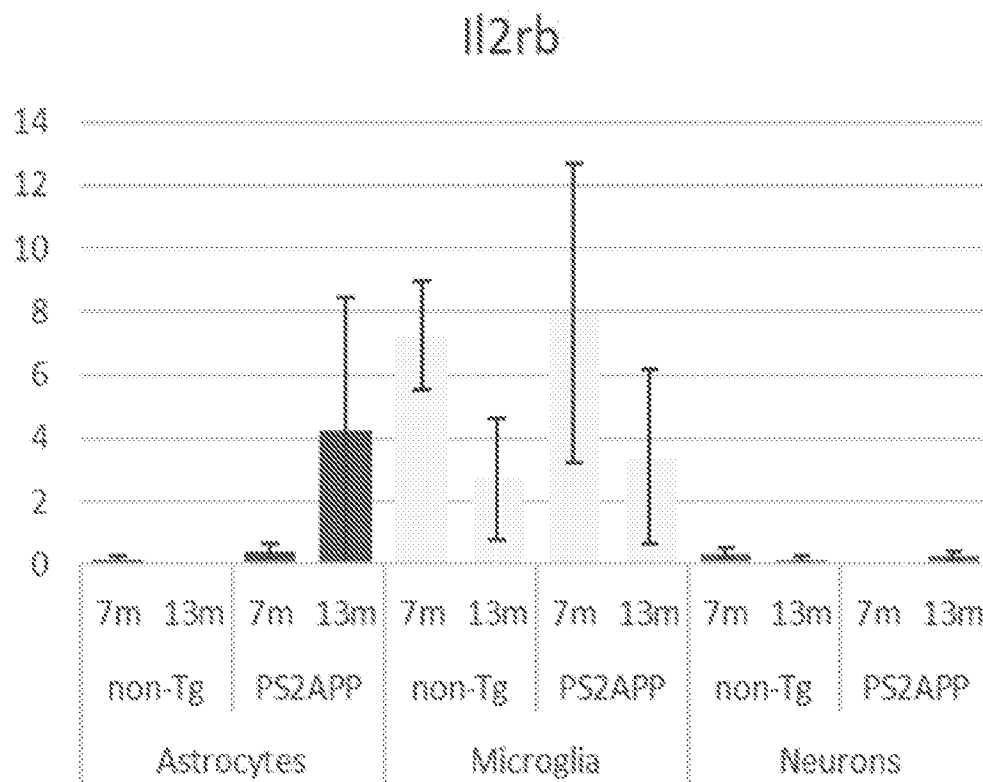
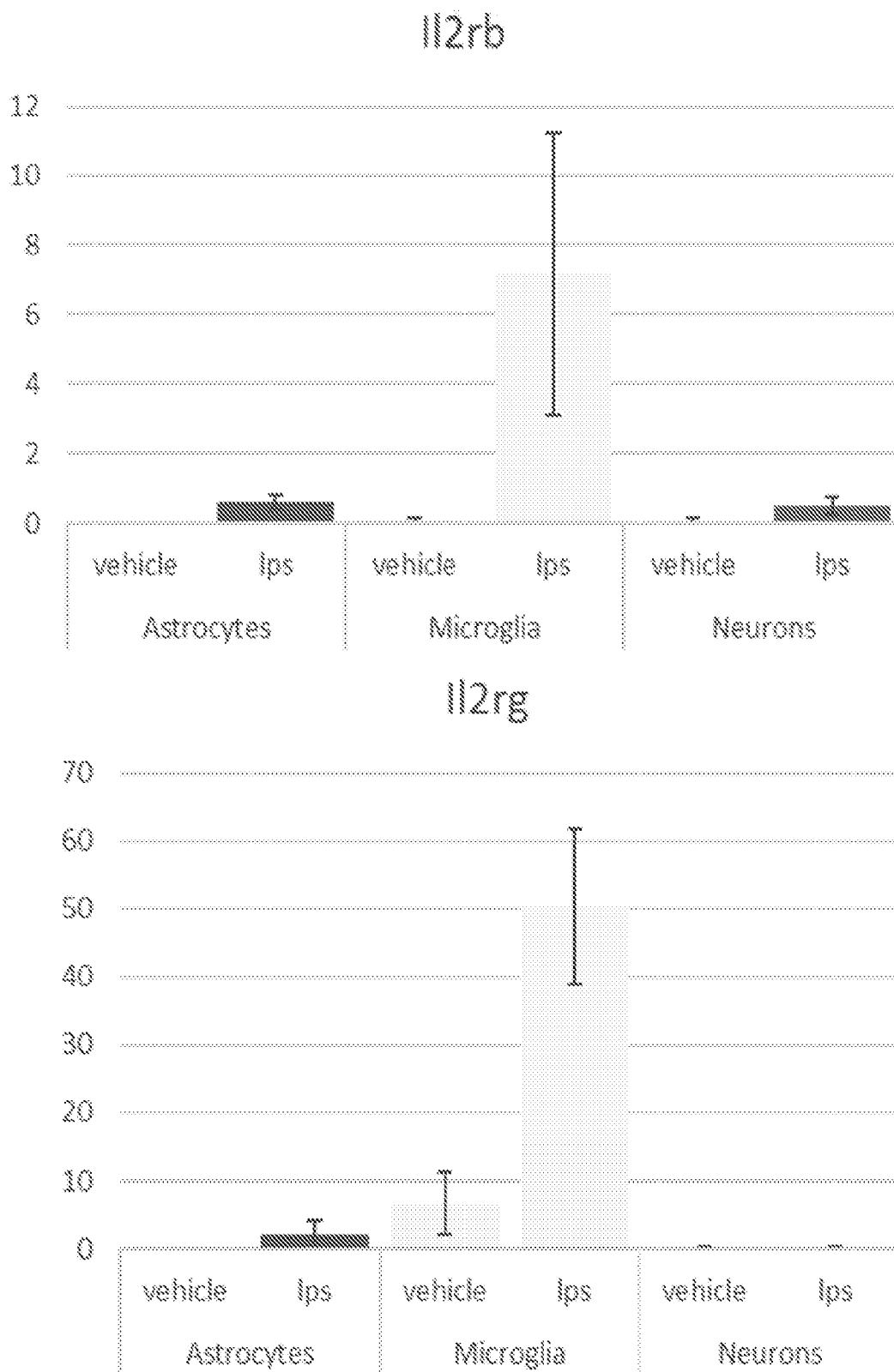


Fig. 39 cont.

**Fig. 40**

**Fig. 40 cont.**

**Fig. 40 cont.**

CHR	BP	SNP	A1	P	Bonf. Correc. P	BETA	Gene(s)
10	6111961	rs35816671	C	1.64E-09	8.03E-03	-0.0579	IL2RA/IL15RA
10	6110925	rs62626317	G	1.18E-08	5.78E-02	-0.0635	IL2RA/IL15RA
10	6107265	rs3118475	C	1.40E-08	6.85E-02	-0.0633	IL2RA/IL15RA
10	6111622	rs41295055	T	1.40E-08	6.85E-02	-0.0633	IL2RA/IL15RA
10	6102012	rs12722489	T	1.47E-08	7.21E-02	-0.0688	IL2RA/IL15RA
10	6118262	rs41295077	A	1.57E-08	7.70E-02	-0.0626	IL2RA/IL15RA
10	6119154	rs41295079	T	1.57E-08	7.70E-02	-0.0626	IL2RA/IL15RA
10	6121285	rs62626325	A	1.57E-08	7.70E-02	-0.0626	IL2RA/IL15RA
10	6122674	rs41295105	G	1.57E-08	7.70E-02	-0.0626	IL2RA/IL15RA
10	6125322	rs41295117	C	1.63E-08	7.96E-02	-0.0666	IL2RA/IL15RA
10	6090322	rs791590	T	1.99E-08	9.74E-02	-0.0675	IL2RA/IL15RA
10	6092210	rs2246031	T	1.99E-08	9.74E-02	-0.0675	IL2RA/IL15RA
10	6075359	rs11256416	T	2.60E-08	1.27E-01	-0.0561	IL2RA/IL15RA
10	6047944	rs764851	T	7.93E-08	3.88E-01	-0.0638	IL2RA/IL15RA
10	6049797	rs41294671	T	7.93E-08	3.88E-01	-0.0638	IL2RA/IL15RA
10	6051741	rs11591302	G	7.93E-08	3.88E-01	-0.0638	IL2RA/IL15RA
10	6080354	rs41294713	C	8.41E-08	4.11E-01	-0.0641	IL2RA/IL15RA
10	6099045	rs2104286	C	1.02E-07	5.00E-01	-0.0537	IL2RA/IL15RA
10	6081230	rs12722515	A	1.09E-07	5.34E-01	-0.0636	IL2RA/IL15RA
10	6058735	rs4747837	G	1.27E-07	6.19E-01	-0.0626	IL2RA/IL15RA

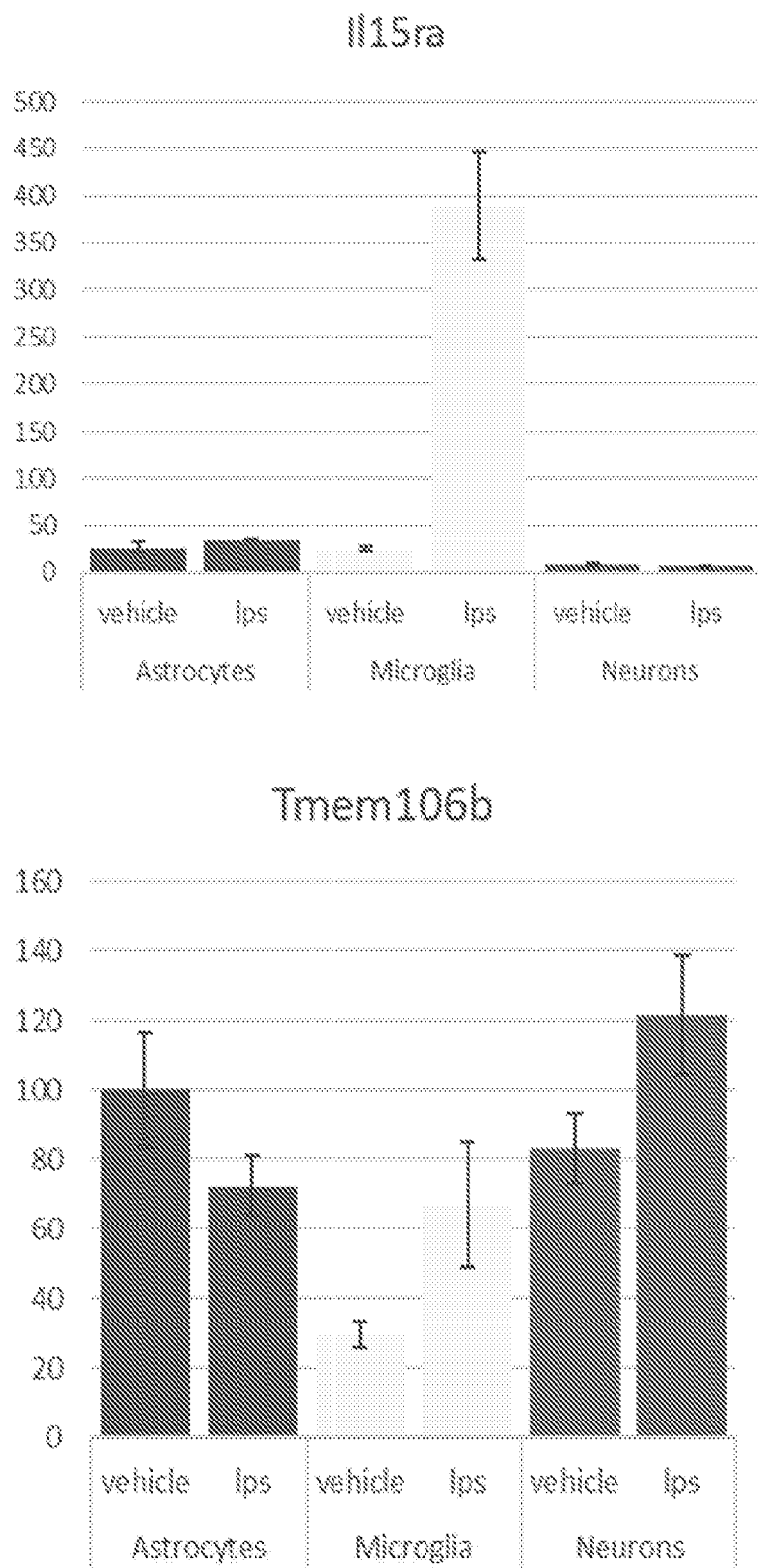
Fig. 41

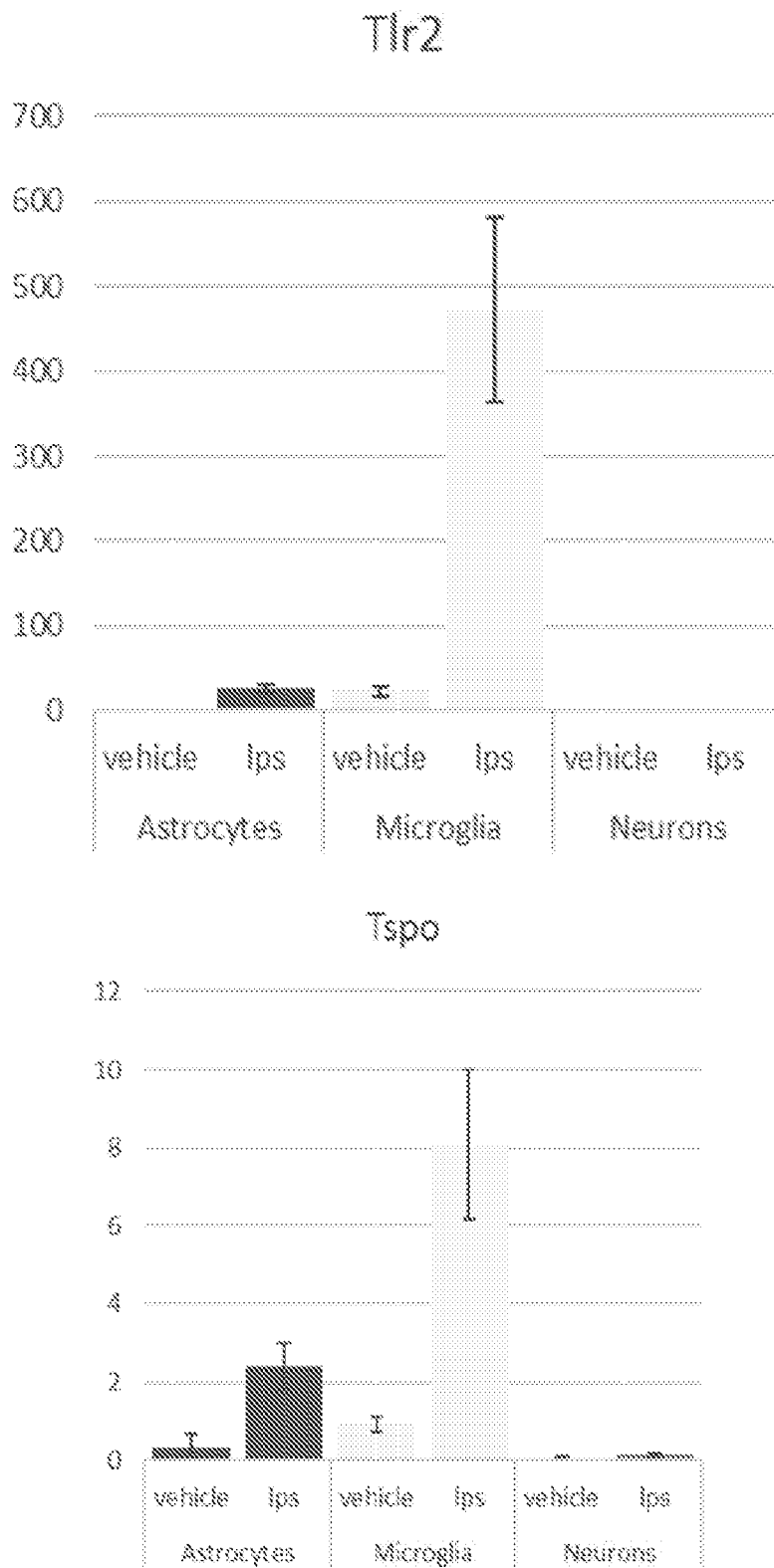
10	6077328	rs12722527	T	1.31E-07	6.42E-01	-0.0626	IL2RA/IL15RA
10	6078390	rs12722523	A	1.31E-07	6.42E-01	-0.0626	IL2RA/IL15RA
10	6076405	rs34795979	A	1.85E-07	9.04E-01	-0.061	IL2RA/IL15RA
10	6071379	rs12722551	T	2.15E-07	1.00E+00	-0.06	IL2RA/IL15RA
10	6070273	rs12722559	A	2.52E-07	1.00E+00	-0.0604	IL2RA/IL15RA
10	6069340	rs12722564	C	2.82E-07	1.00E+00	-0.0597	IL2RA/IL15RA
10	6069893	rs12722561	T	2.82E-07	1.00E+00	-0.0597	IL2RA/IL15RA
10	6071284	rs12722553	A	2.82E-07	1.00E+00	-0.0597	IL2RA/IL15RA
10	6072504	rs4625363	G	2.82E-07	1.00E+00	-0.0597	IL2RA/IL15RA
10	6073481	rs200481846	A	2.82E-07	1.00E+00	-0.0597	IL2RA/IL15RA
10	6038478	rs17147986	A	3.78E-07	1.00E+00	-0.0602	IL2RA/IL15RA
10	6069288	rs201247027	C	4.67E-07	1.00E+00	-0.0588	IL2RA/IL15RA
10	6129874	rs7898890	A	5.51E-07	1.00E+00	-0.0442	IL2RA/IL15RA
10	6102726	rs12722488	TA	5.76E-07	1.00E+00	-0.0506	IL2RA/IL15RA
14	77671804	rs26555996	T	1.79E-06	1.00E+00	0.0405	TMEM63C
14	77667213	rs7140198	G	1.81E-06	1.00E+00	0.0406	TMEM63C
10	6049344	rs10795737	T	2.11E-06	1.00E+00	-0.0465	RBM17
18	31799879	rs72961236	G	2.21E-06	1.00E+00	-0.0577	NOL4
6	52627698	rs2608632	T	2.90E-06	1.00E+00	0.0412	GSTA1/GSTA2/GSTA7P/GSTA5
6	52627710	rs2749017	G	2.90E-06	1.00E+00	0.0412	GSTA1/GSTA2/GSTA7P/GSTA5
1	118141596	rs10923350	G	2.94E-06	1.00E+00	0.0472	FAM46C
1	118147142	rs1992117	G	2.94E-06	1.00E+00	0.0472	FAM46C

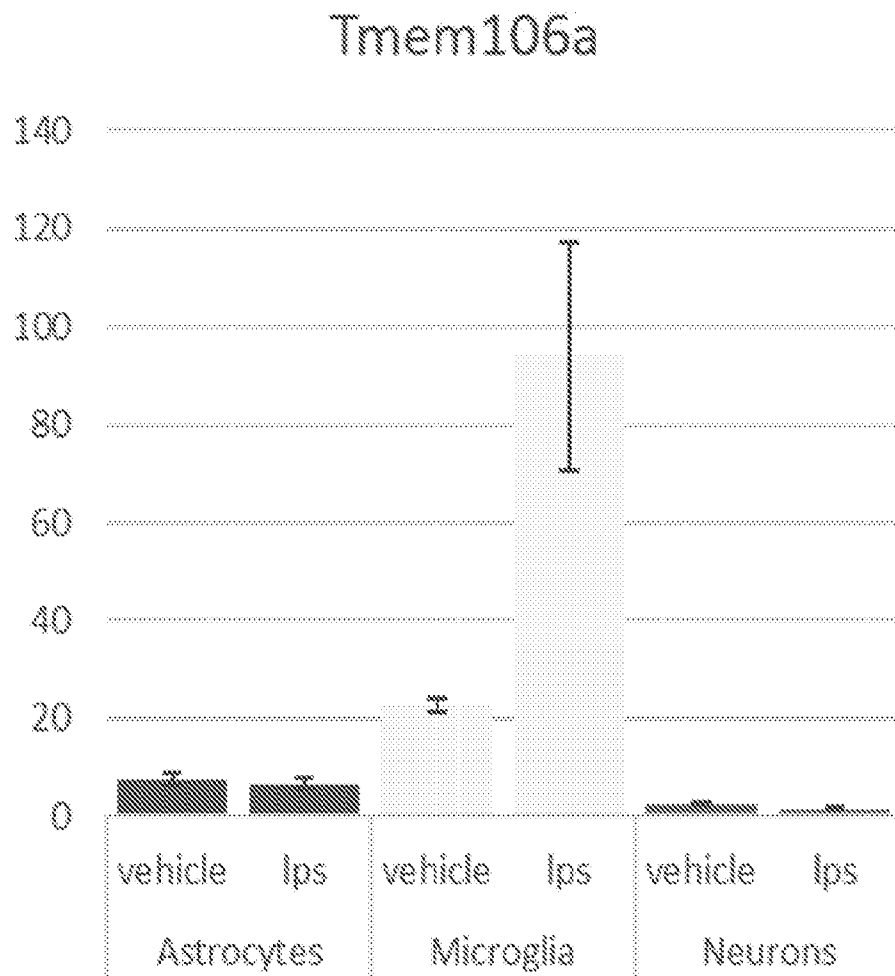
Fig. 41 cont.

1	118151846	rs893230	T	2.94E-06	1.00E+00	0.0472	FAM46C
1	118151935	rs1808809	T	2.94E-06	1.00E+00	0.0472	FAM46C
6	52620740	rs2474462	A	2.99E-06	1.00E+00	0.041	GSTA1/GSTA2/GSTA7P/GSTA5
6	52623807	rs2144698	T	2.99E-06	1.00E+00	0.041	GSTA1/GSTA2/GSTA7P/GSTA5
1	118152241	rs10923354	C	3.15E-06	1.00E+00	0.047	FAM46C
1	118154831	rs3767809	C	3.15E-06	1.00E+00	0.047	FAM46C
1	118154938	rs3767810	G	3.15E-06	1.00E+00	0.047	FAM46C
1	118155145	rs143811622	A	3.15E-06	1.00E+00	0.047	FAM46C
1	118155152	rs199858224	G	3.15E-06	1.00E+00	0.047	FAM46C
10	6049417	rs10795738	C	3.62E-06	1.00E+00	-0.0458	R8M17
6	52675433	rs11752938	T	4.03E-06	1.00E+00	0.0389	GSTA1/GSTA2/GSTA7P/GSTA5
6	52695288	rs12529615	A	4.03E-06	1.00E+00	0.0389	GSTA1/GSTA2/GSTA7P/GSTA5
6	52695461	rs10948726	T	4.03E-06	1.00E+00	0.0389	GSTA1/GSTA2/GSTA7P/GSTA5
11	112068219	rs11214126	T	4.65E-06	1.00E+00	-0.0374	IL18/BCO2/TEX12/PTS
5	35702209	rs35297265	G	4.66E-06	1.00E+00	-0.0389	SPEF2
5	35702212	rs11357084	GA	4.66E-06	1.00E+00	-0.0389	SPEF2
5	35702731	rs12514664	G	4.66E-06	1.00E+00	-0.0389	SPEF2
6	52632479	rs2749033	A	4.68E-06	1.00E+00	0.0401	GSTA1/GSTA2/GSTA7P/GSTA5
6	166518499	rs35711971	CCA	4.69E-06	1.00E+00	-0.0427	GSTA1/GSTA2/GSTA7P/GSTA5
2	191284865	rs72903383	G	4.73E-06	1.00E+00	-0.0485	INPP1/MFSD6
6	166517329	rs1445287	T	4.80E-06	1.00E+00	-0.0427	GSTA1/GSTA2/GSTA7P/GSTA5

Fig. 41 cont.

**Fig. 42**

**Fig. 42 cont.**

**Fig. 42 cont.**

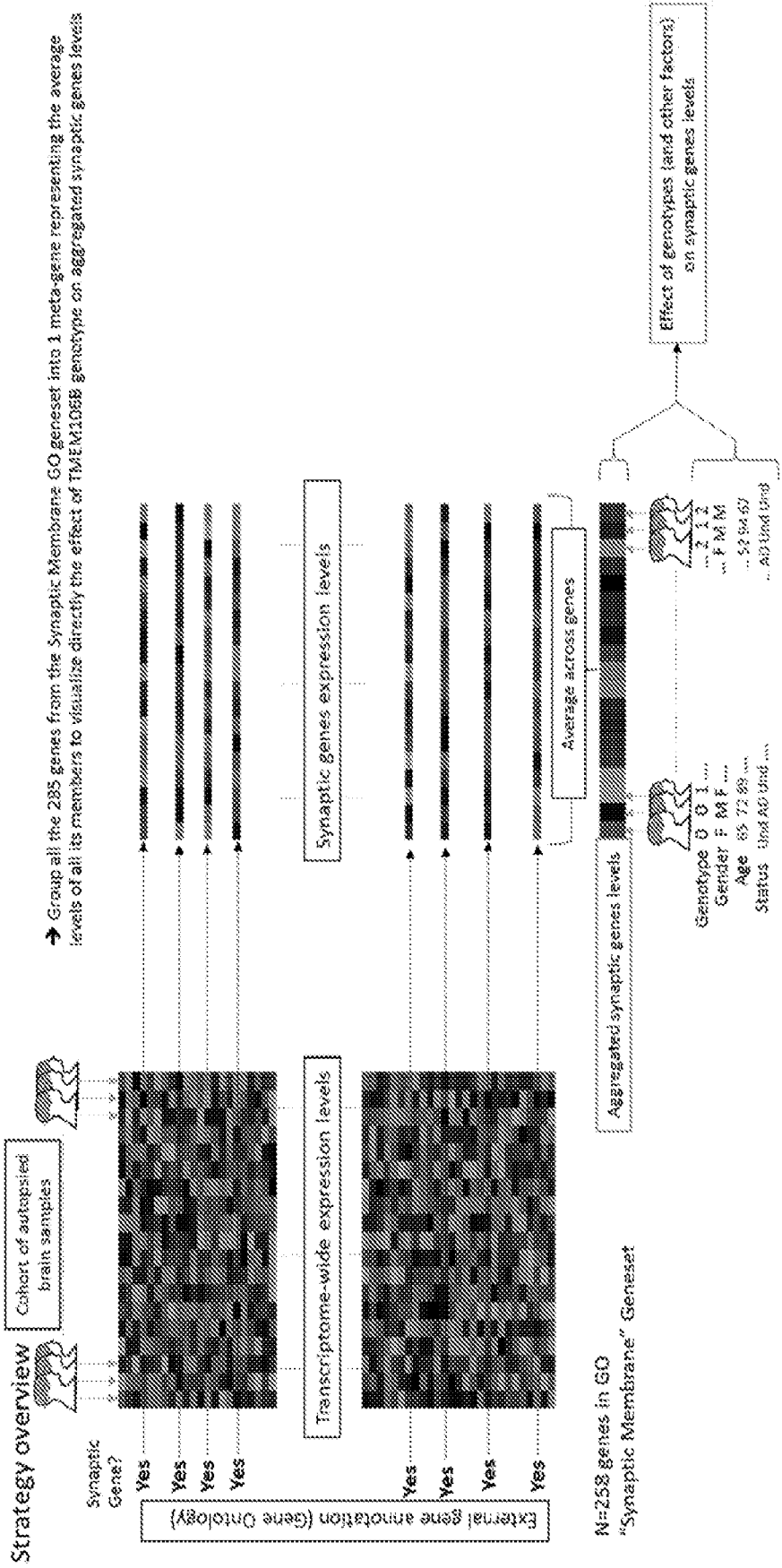
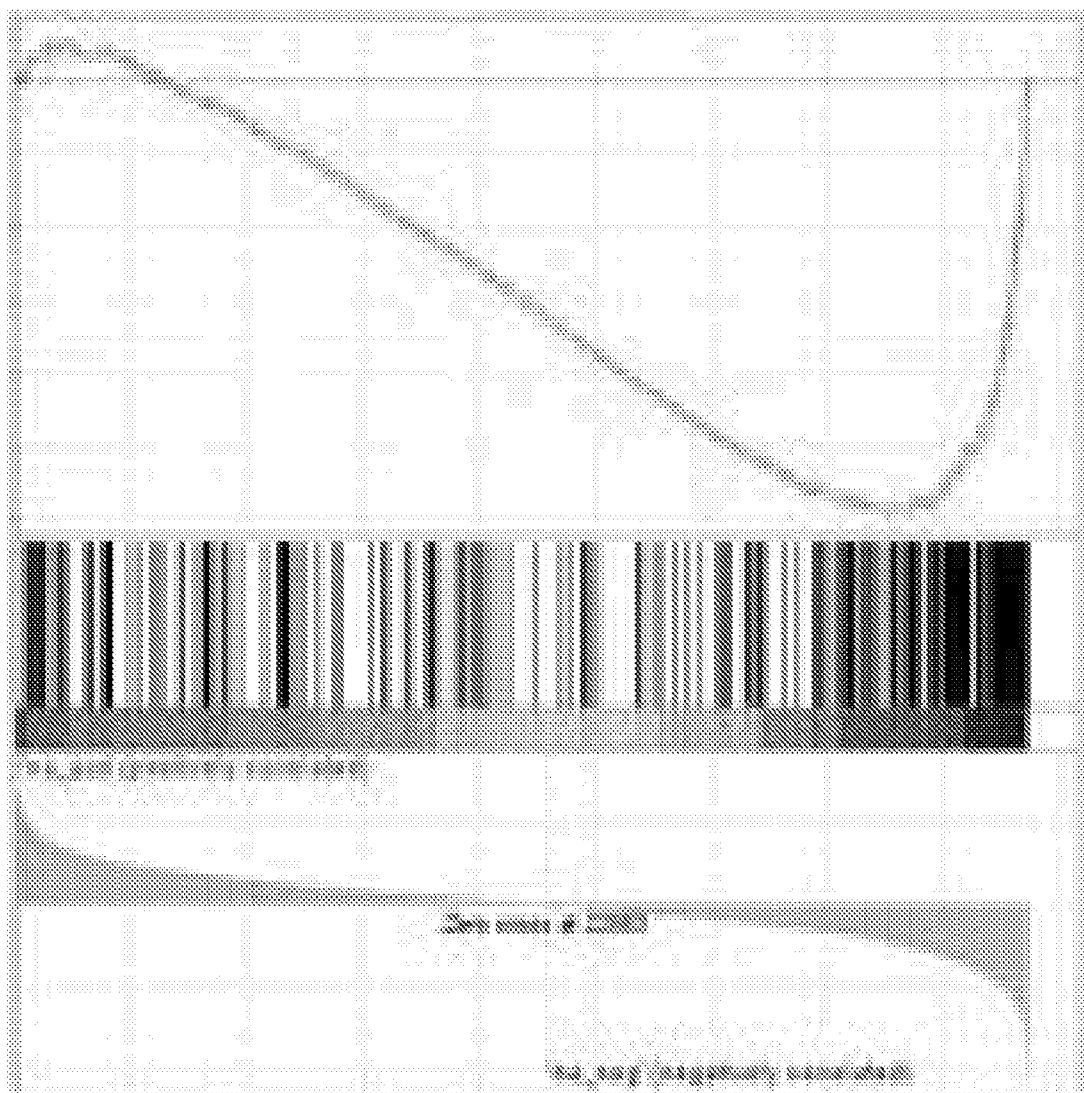
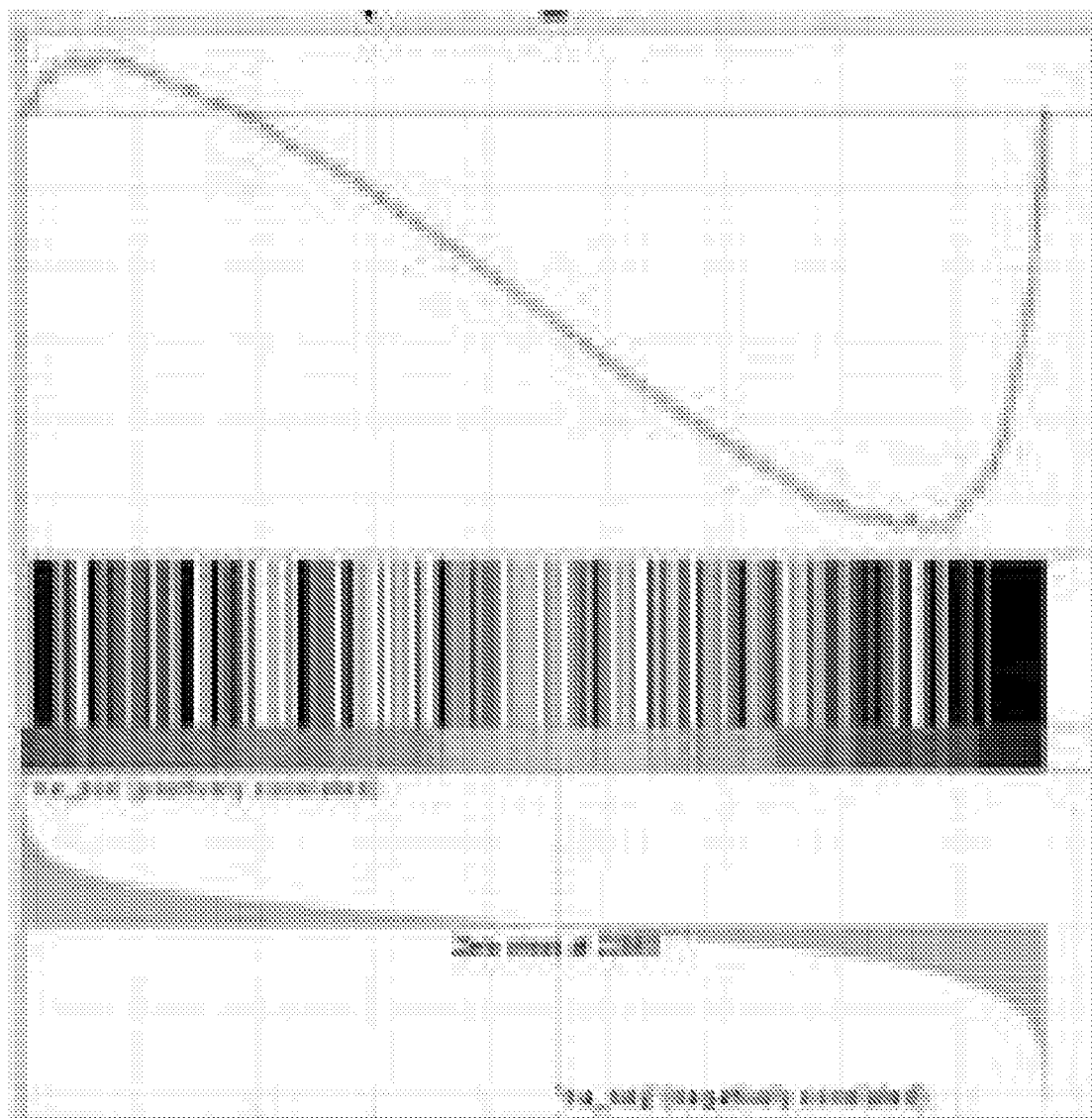
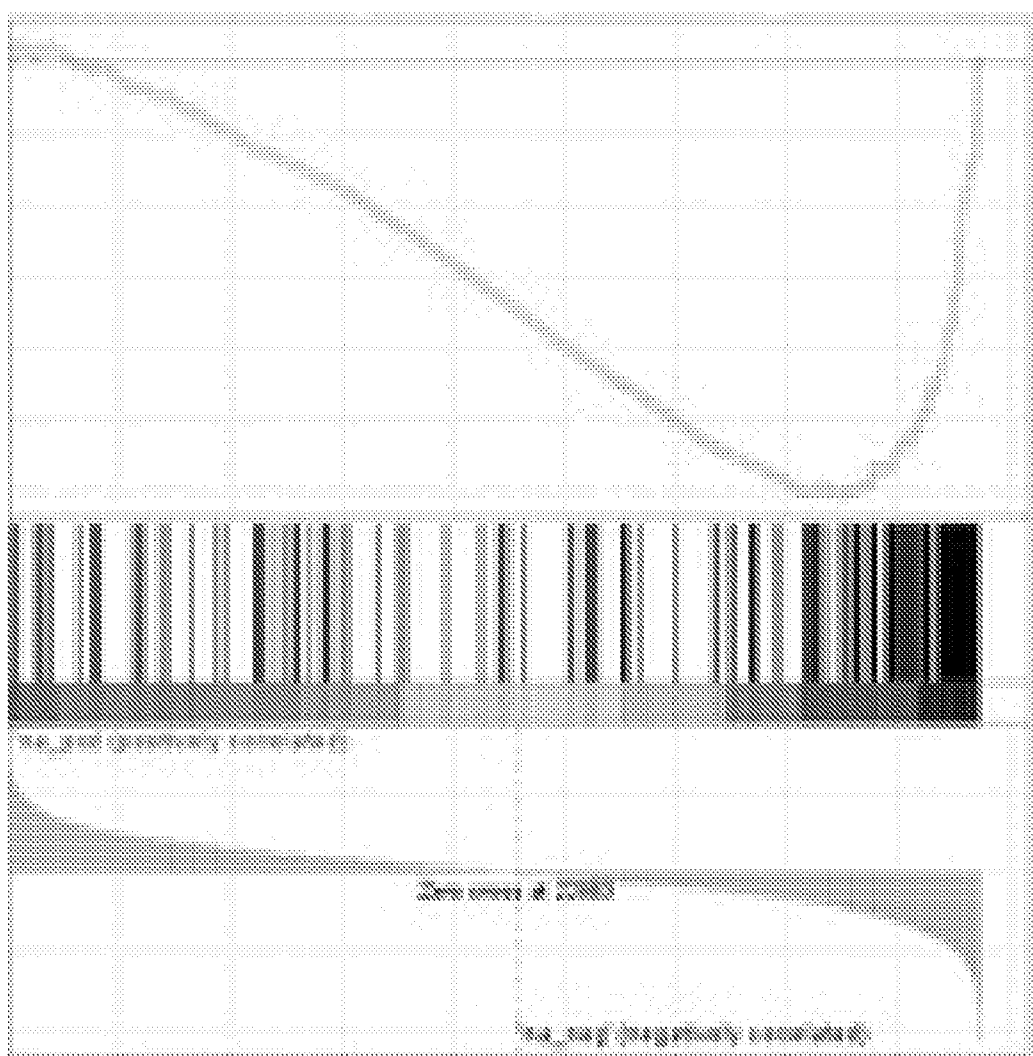
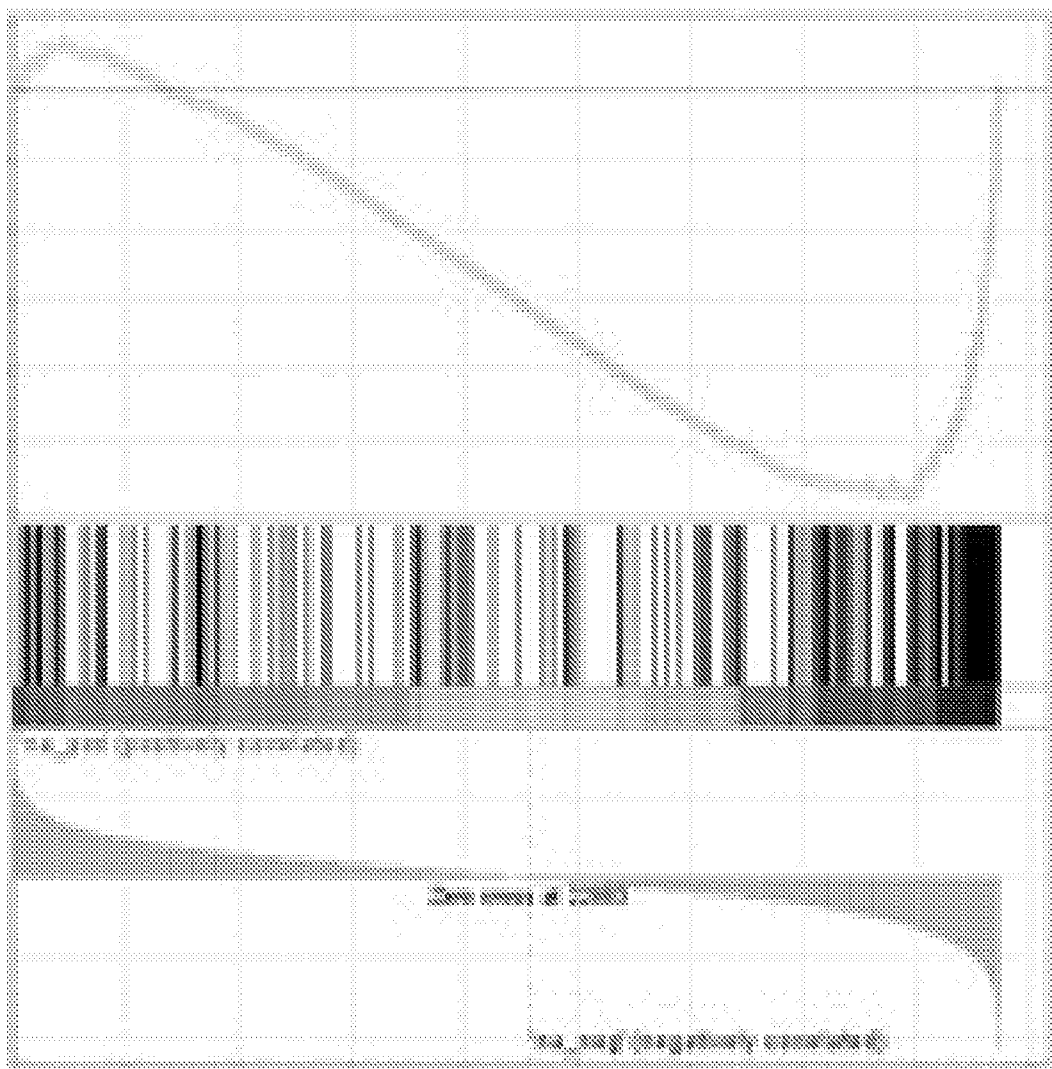


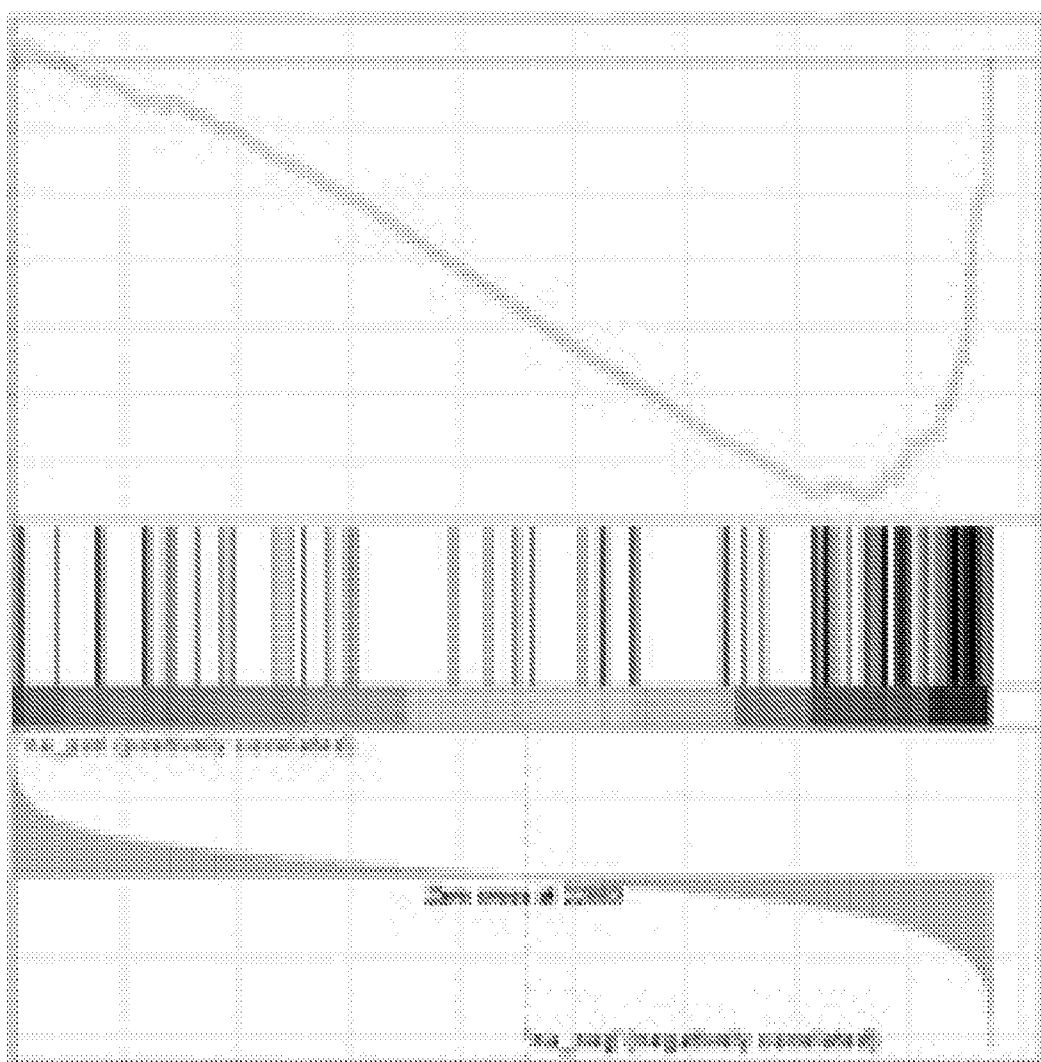
Fig. 43

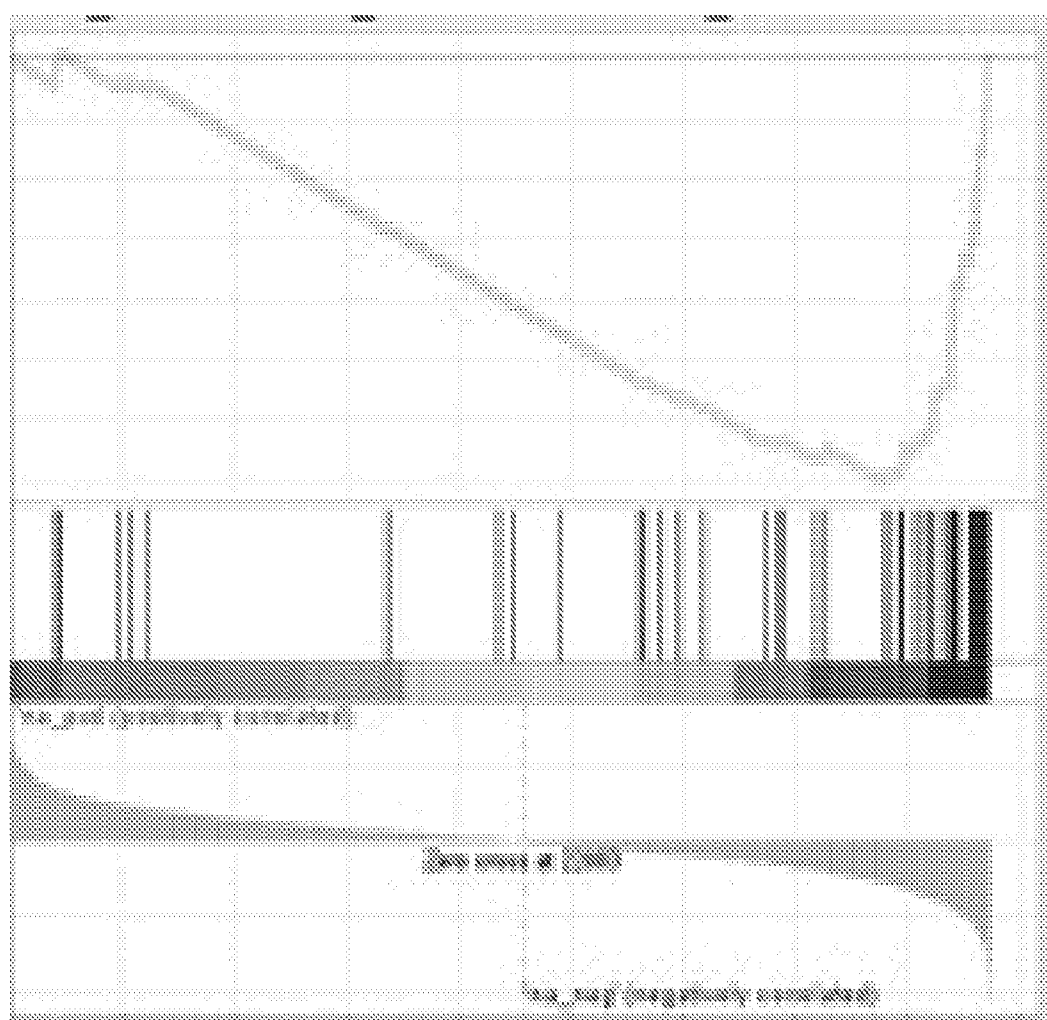
**Fig. 44A**

**Fig. 44B**

**Fig. 44C**

**Fig. 44D**

**Fig. 44E**

**Fig. 44F**

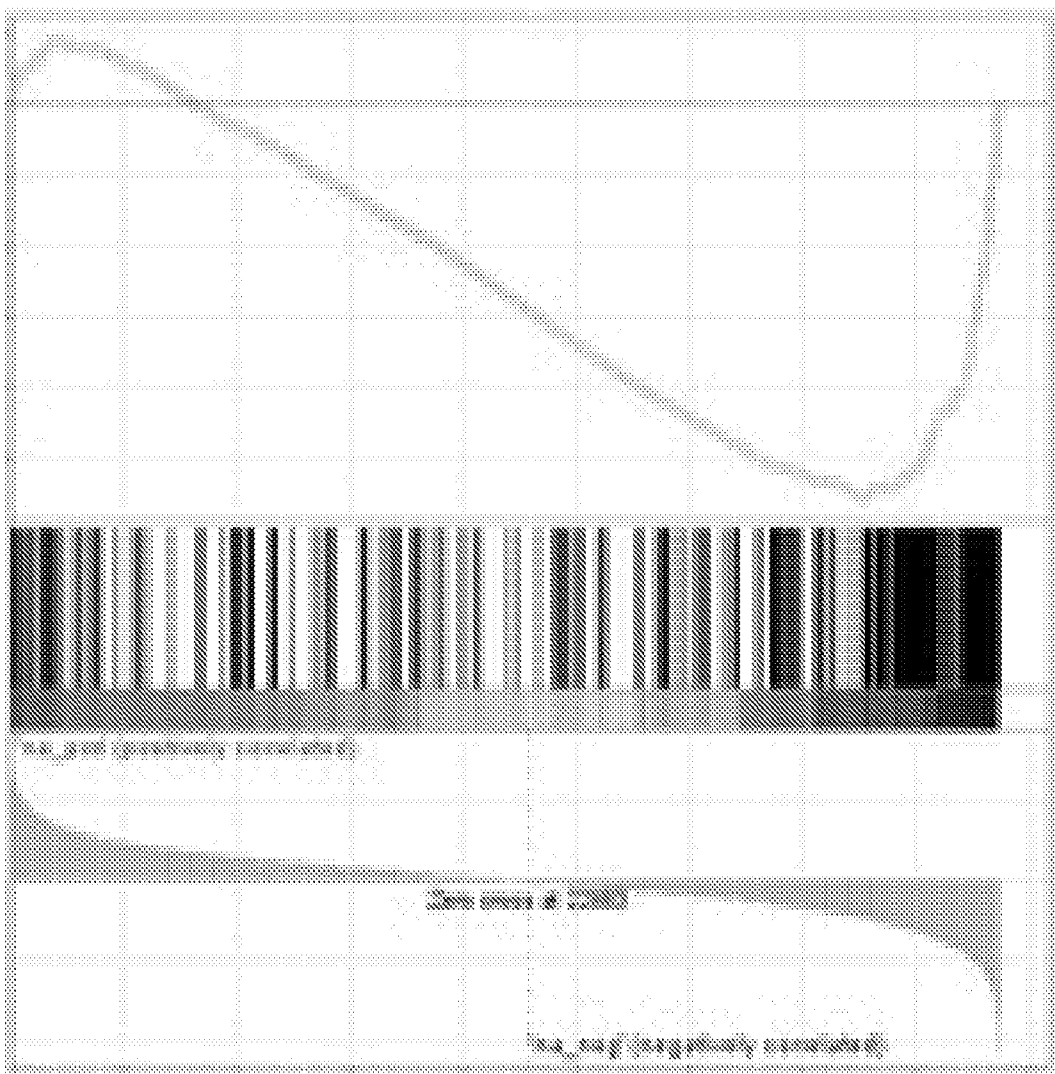
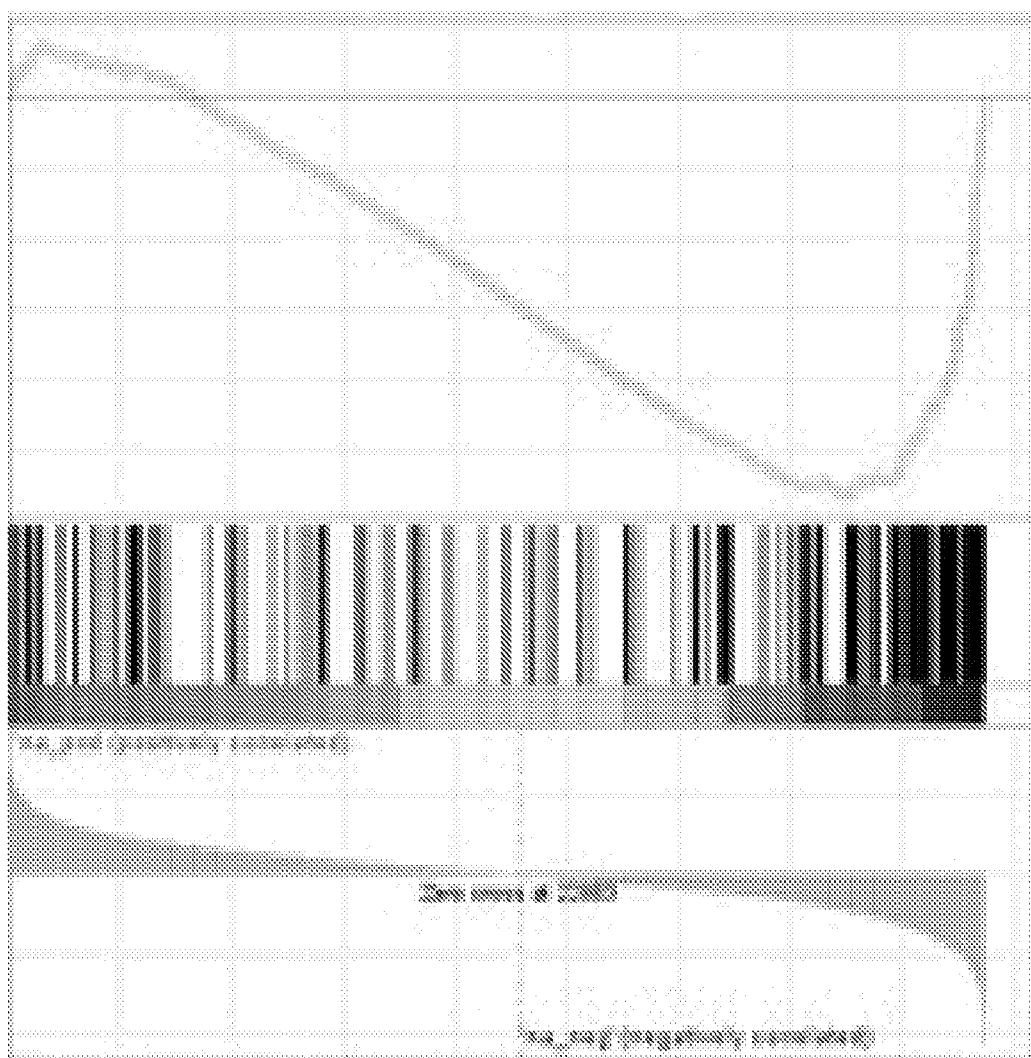
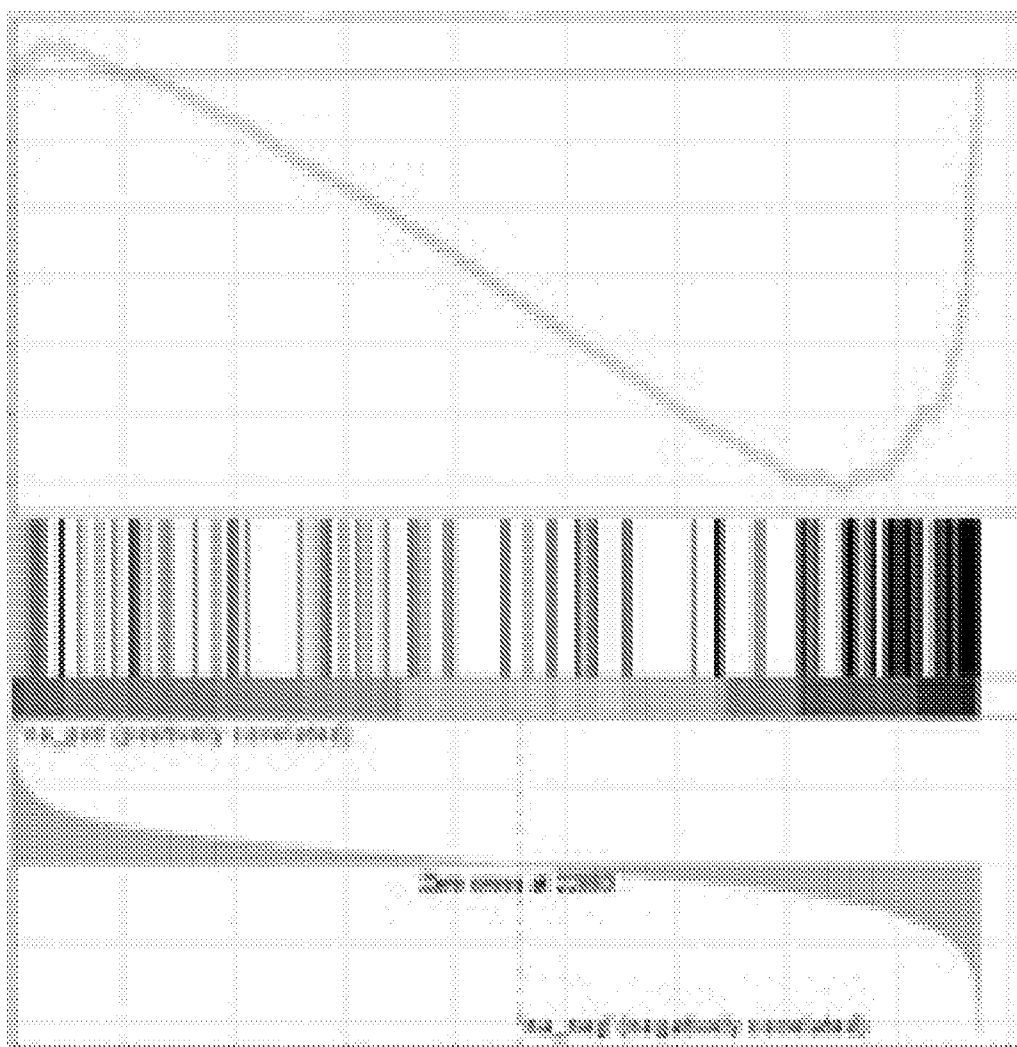
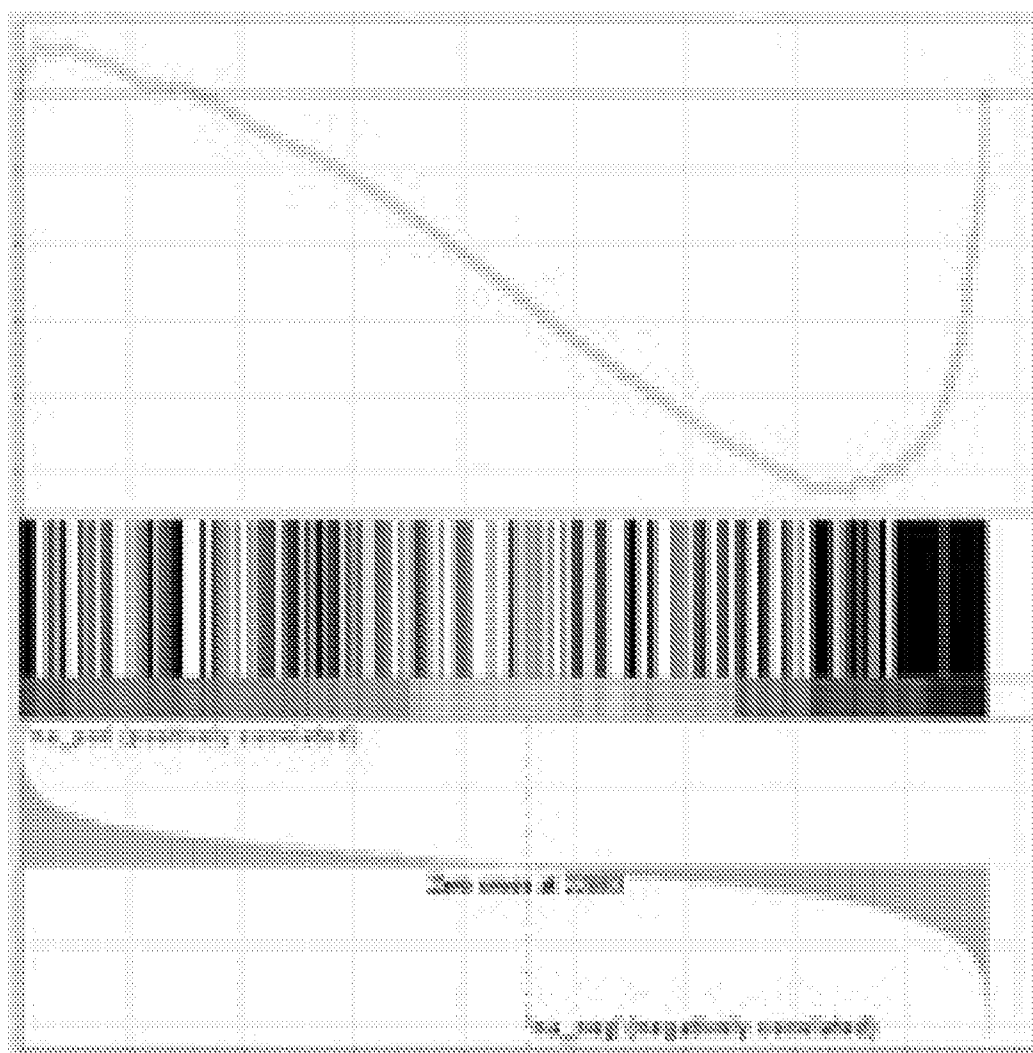
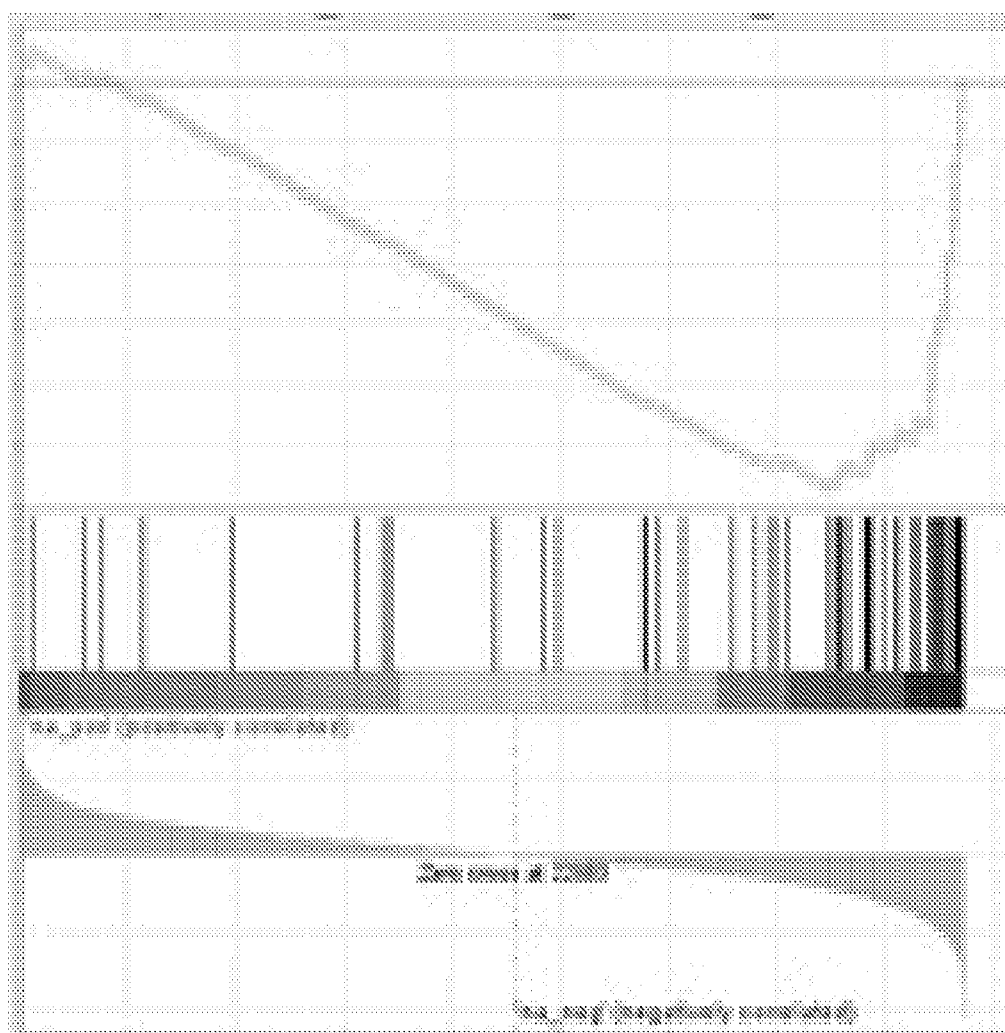


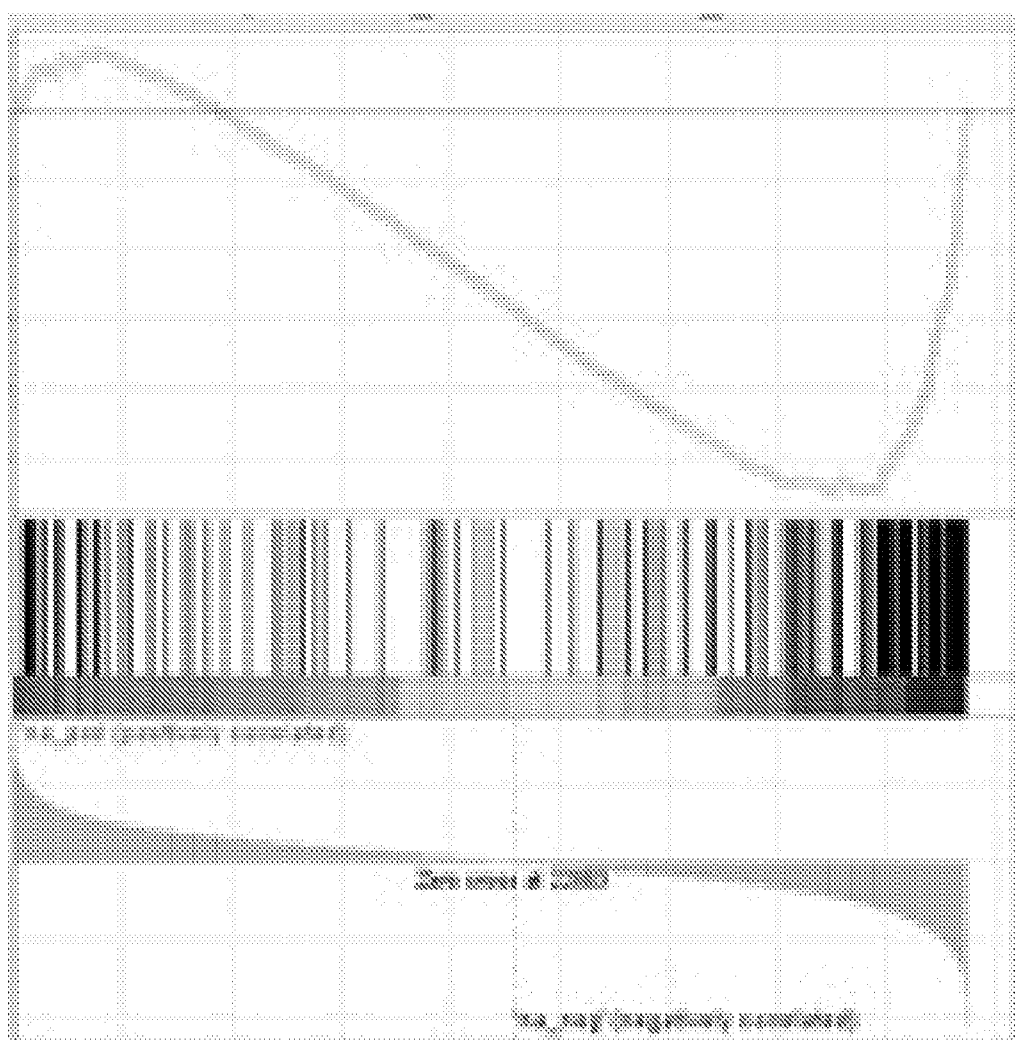
Fig. 44G

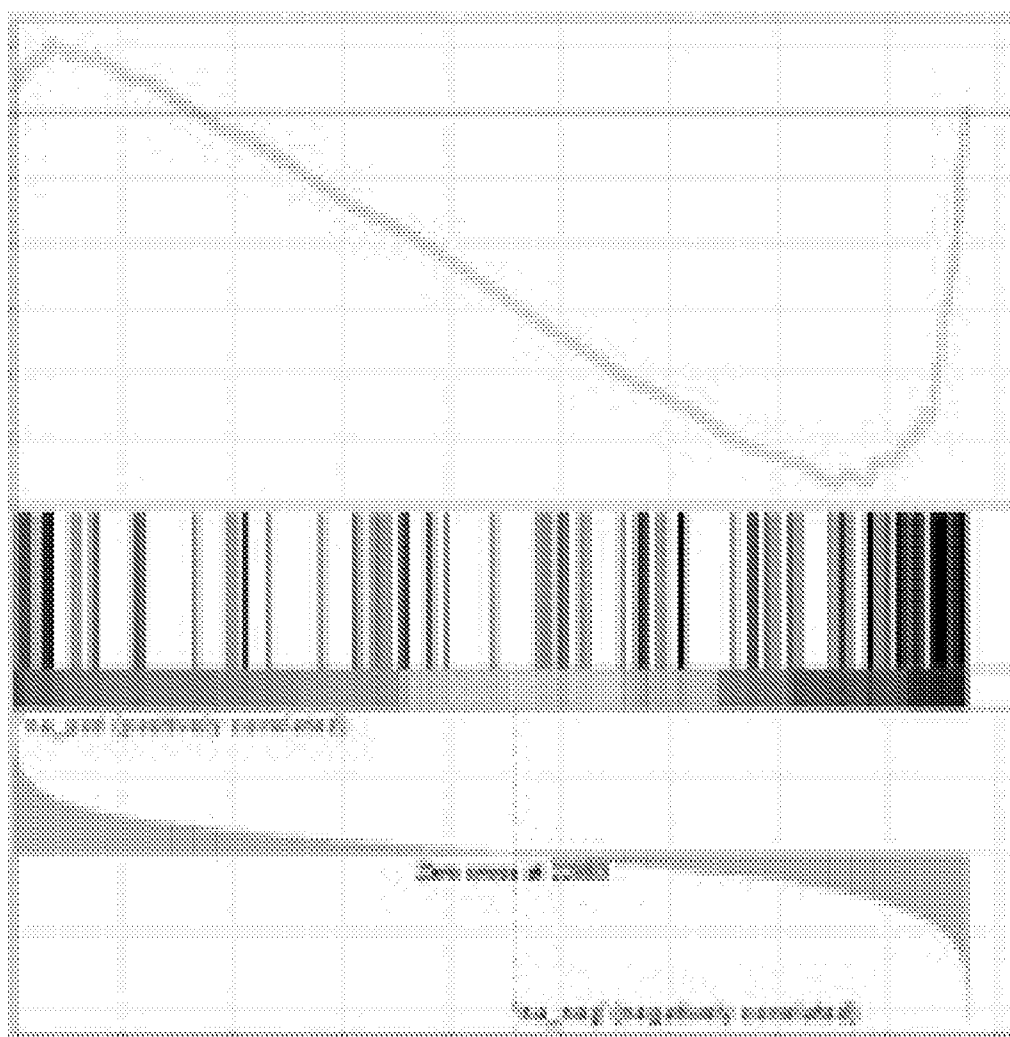
**Fig. 44H**

**Fig. 44I**

**Fig. 44J**

**Fig. 44K**

**Fig. 44L**

**Fig. 44M**

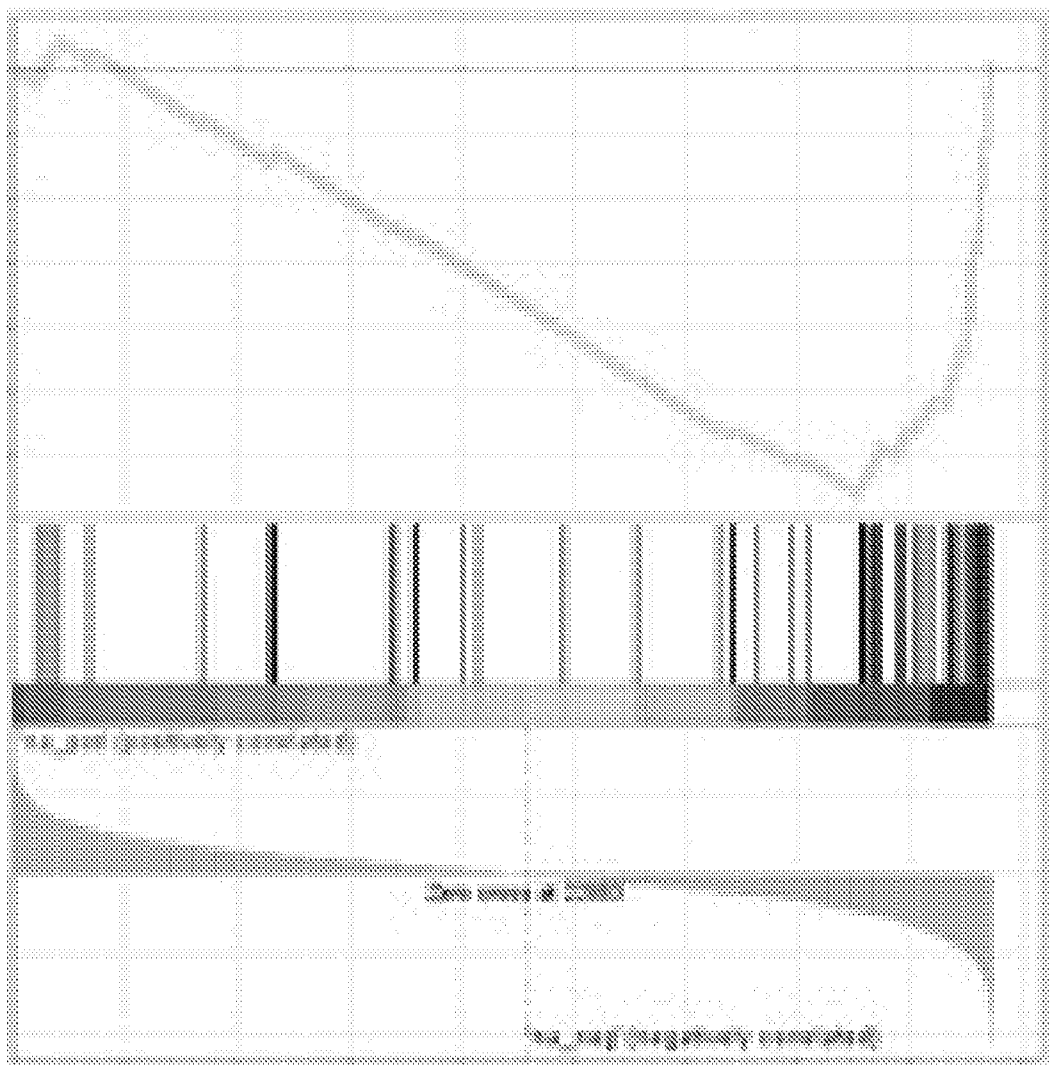
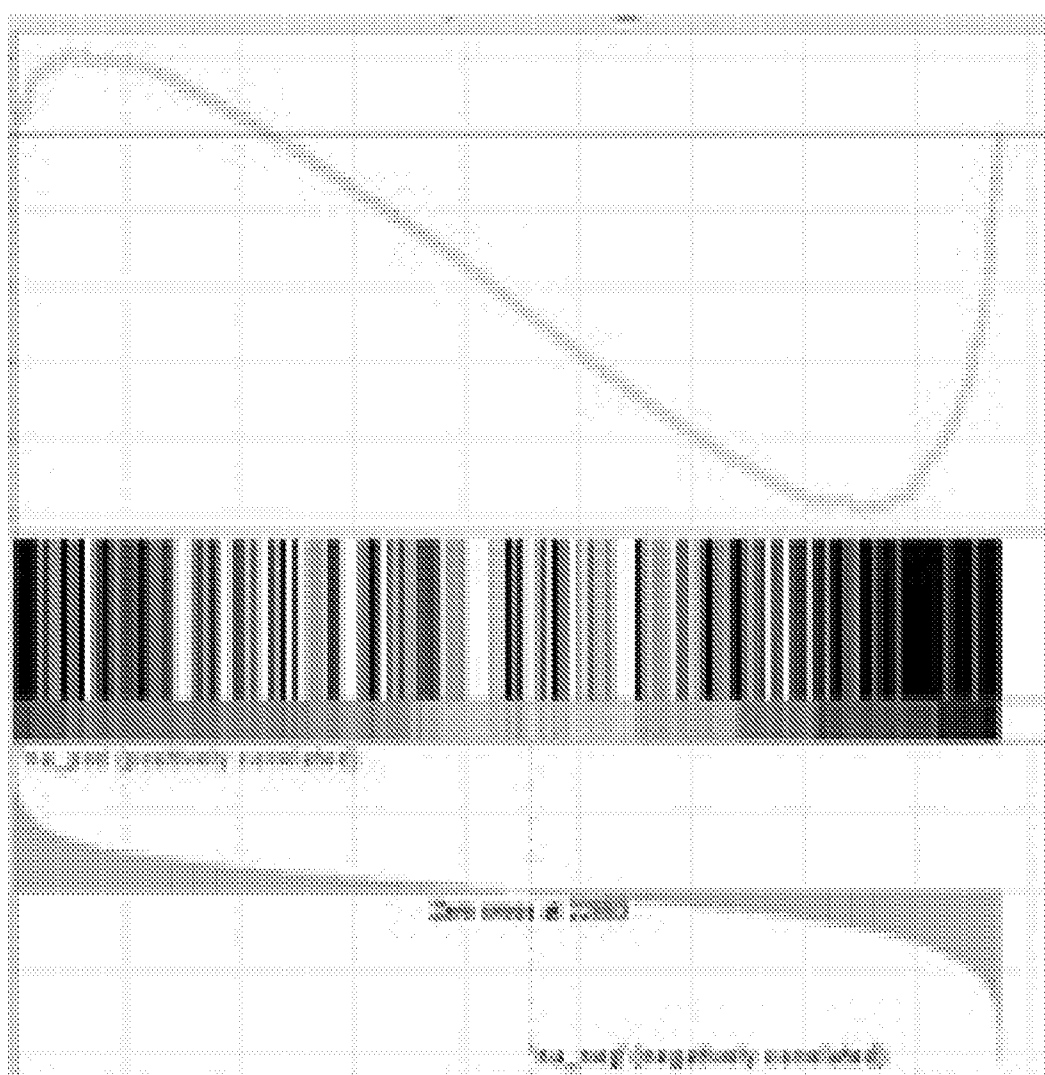
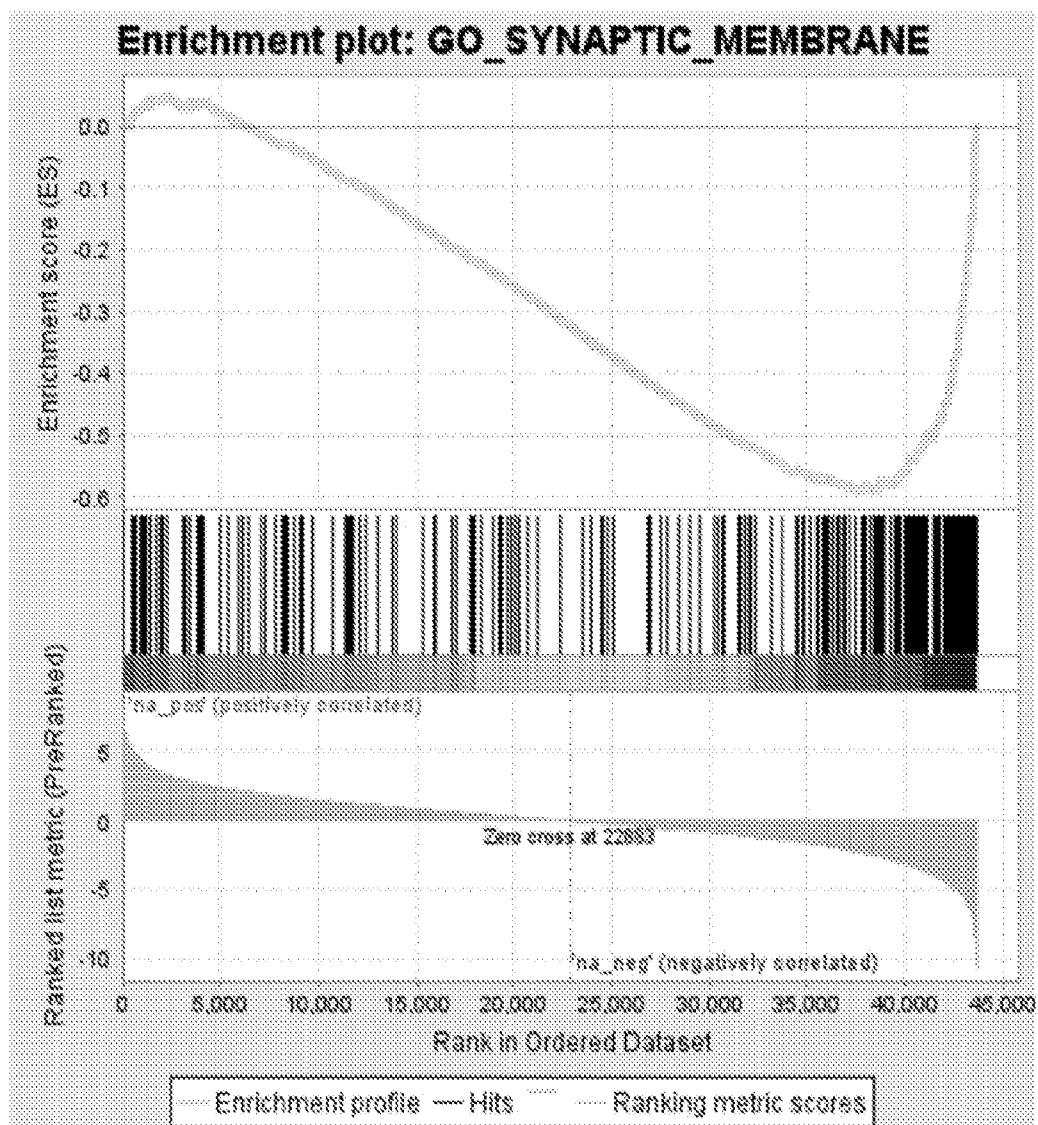


Fig. 44N

**Fig. 440**

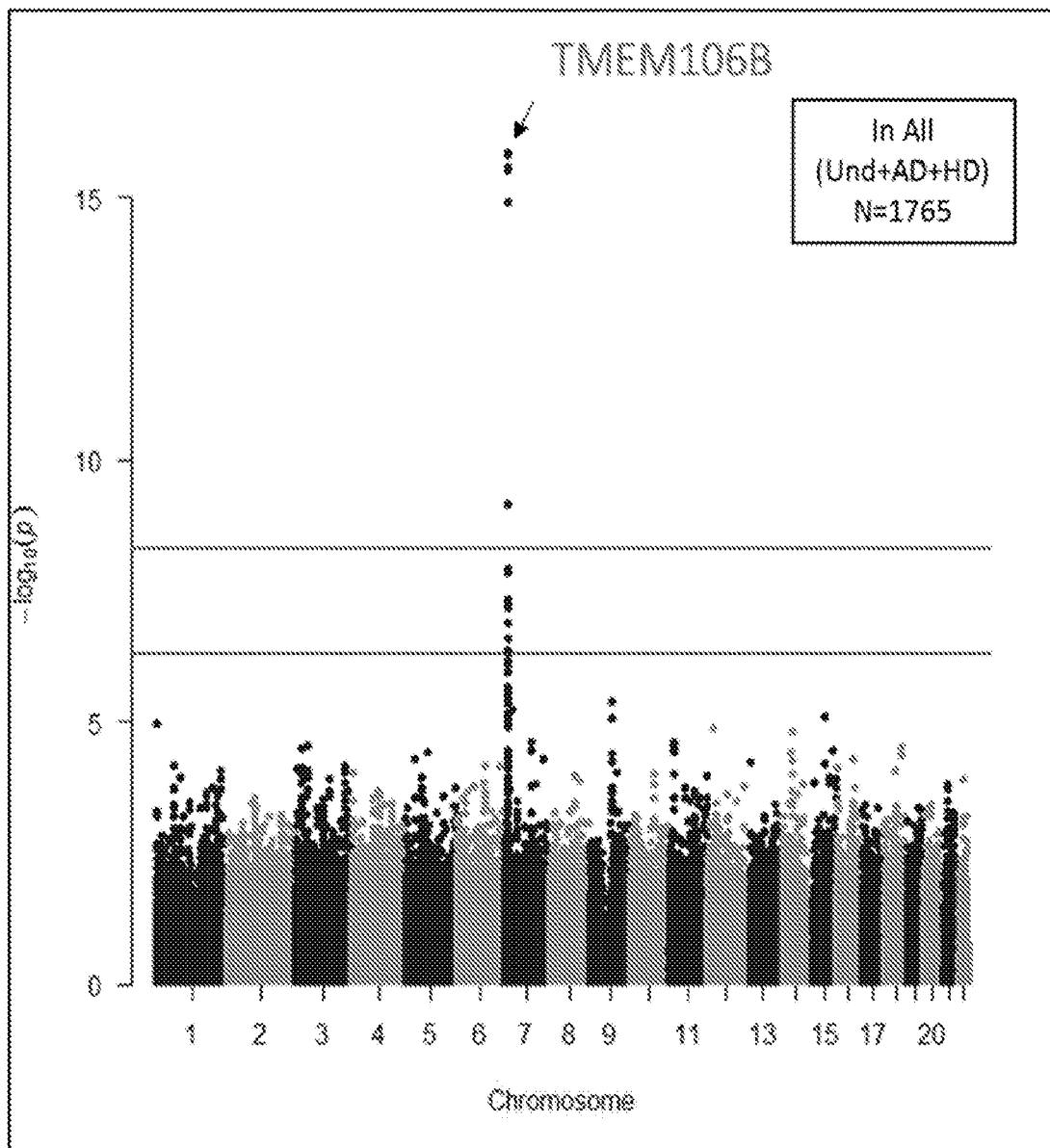
**Fig. 45**

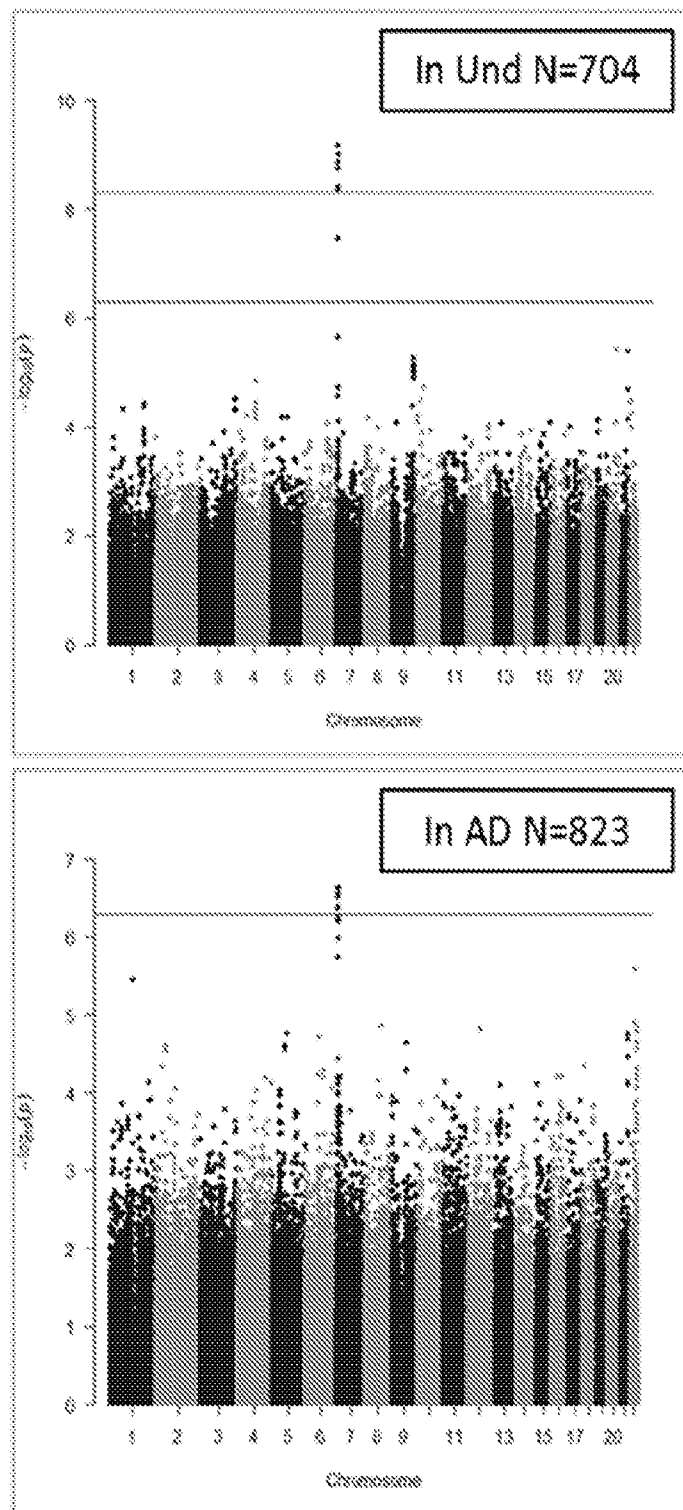
211	GRIA4				42885	-5.029	-0.3818	Yes
212	LRSTM2				42877	-5.047	-0.3860	Yes
213	KOHL11				42880	-5.053	-0.3489	Yes
214	SNAP25				42890	-5.073	-0.3441	Yes
215	LRFM3				42762	-5.163	-0.3393	Yes
216	PDE2A				42756	-5.170	-0.3331	Yes
217	GABRB2				42835	-5.288	-0.3286	Yes
218	LINTA				42849	-5.311	-0.3226	Yes
219	GAC2				42857	-5.322	-0.3163	Yes
220	SHISA2				42902	-5.402	-0.3108	Yes
221	CDK5				42914	-5.432	-0.3045	Yes
222	GRP1				42932	-5.443	-0.2984	Yes
223	DNAJA3				42962	-5.500	-0.2925	Yes
224	MAGEE1				42972	-5.520	-0.2880	Yes
225	DLG3				42983	-5.539	-0.2798	Yes
226	CLSTN3				43002	-5.573	-0.2733	Yes
227	GRM2				43030	-5.628	-0.2672	Yes
228	PODH8				43050	-5.679	-0.2608	Yes
229	RGS14				43059	-5.690	-0.2542	Yes
230	GRIN2A				43097	-5.777	-0.2481	Yes
231	CHRNA2				43104	-5.787	-0.2413	Yes
232	ABHDAP32				43153	-5.886	-0.2353	Yes
233	GABRA4				43168	-5.919	-0.2285	Yes
234	NRXN1				43172	-5.925	-0.2215	Yes
235	LRSTM1				43220	-6.067	-0.2153	Yes

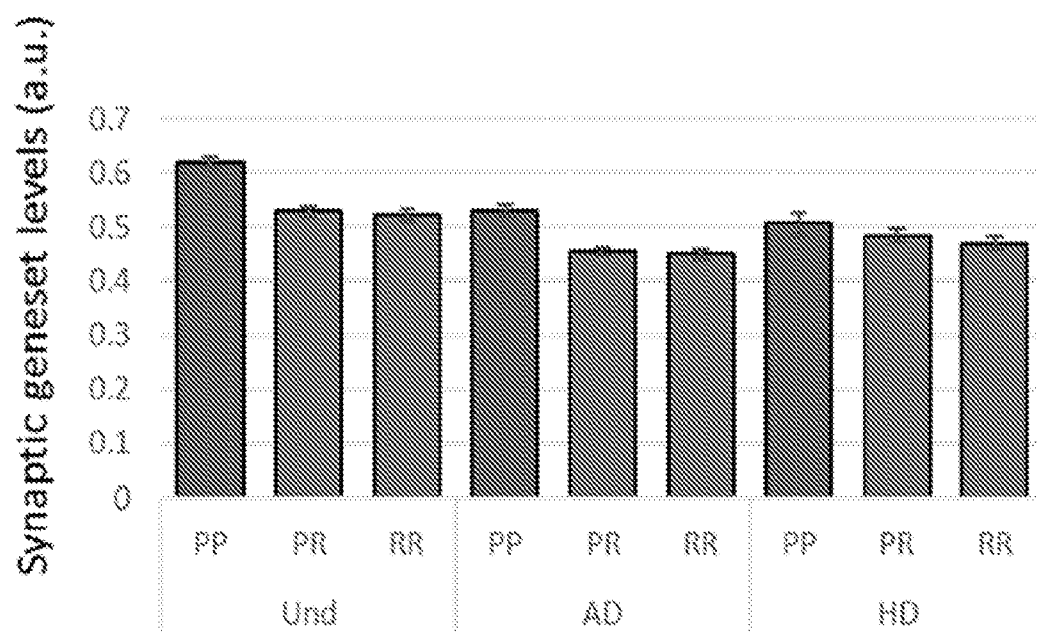
Fig. 46

236	C4QPS2			43224	-6.073	-0.2081	Yes
237	DLG3			43230	-6.107	-0.2008	Yes
238	SYE			43256	-6.154	-0.1940	Yes
239	LPRC4			43287	-6.235	-0.1875	Yes
240	QABR2			43311	-6.311	-0.1802	Yes
241	QABR2			43350	-6.459	-0.1733	Yes
242	QABR2			43352	-6.482	-0.1656	Yes
243	QABR2			43354	-6.471	-0.1578	Yes
244	GLR2			43372	-6.514	-0.1504	Yes
245	PIK2A			43426	-6.739	-0.1435	Yes
246	FLR13			43452	-6.853	-0.1358	Yes
247	AIP2B1			43460	-6.998	-0.1281	Yes
248	CAMK2A			43461	-6.990	-0.1197	Yes
250	CLSTN1			43504	-7.096	-0.1032	Yes
251	KOTD8			43538	-7.358	-0.0952	Yes
252	CBLN1			43558	-7.526	-0.0885	Yes
253	QABR1			43559	-7.530	-0.0773	Yes
254	CHRNA7			43564	-7.722	-0.0681	Yes
255	LPR1B4			43567	-7.729	-0.0589	Yes
256	CAMK2B2			43579	-7.883	-0.0487	Yes
257	PSD3			43582	-7.923	-0.0402	Yes
258	KCM2			43591	-8.039	-0.0307	Yes
259	SEMA4E			43607	-8.363	-0.0210	Yes
260	CHRM3			43613	-8.581	-0.0108	Yes
261	EPHA4			43622	-9.291	0.0901	Yes

Fig. 46 cont.

**Fig. 47**

**Fig. 47cont.**

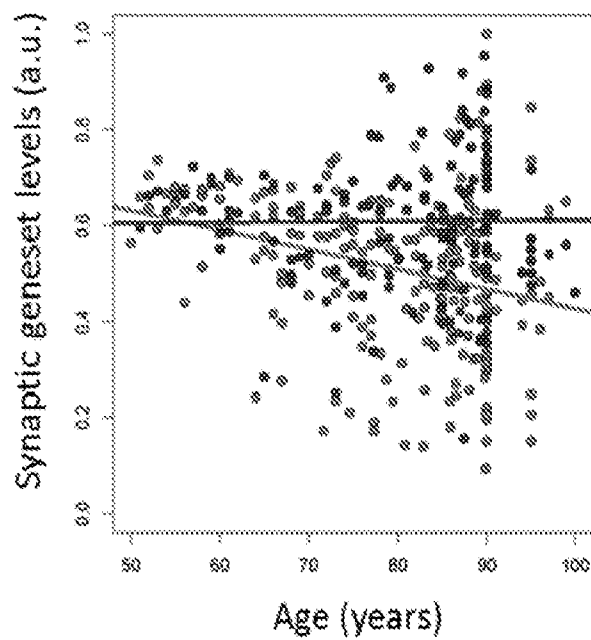


N=147, 340, 217, 179, 381, 263, 31, 77, 62

Means are presented, error bars are SEM

Fig. 48

84/102



TMEM106B PP

TMEM106B RR

N=147 PP (blue), 217 RR (red)

Fig. 49

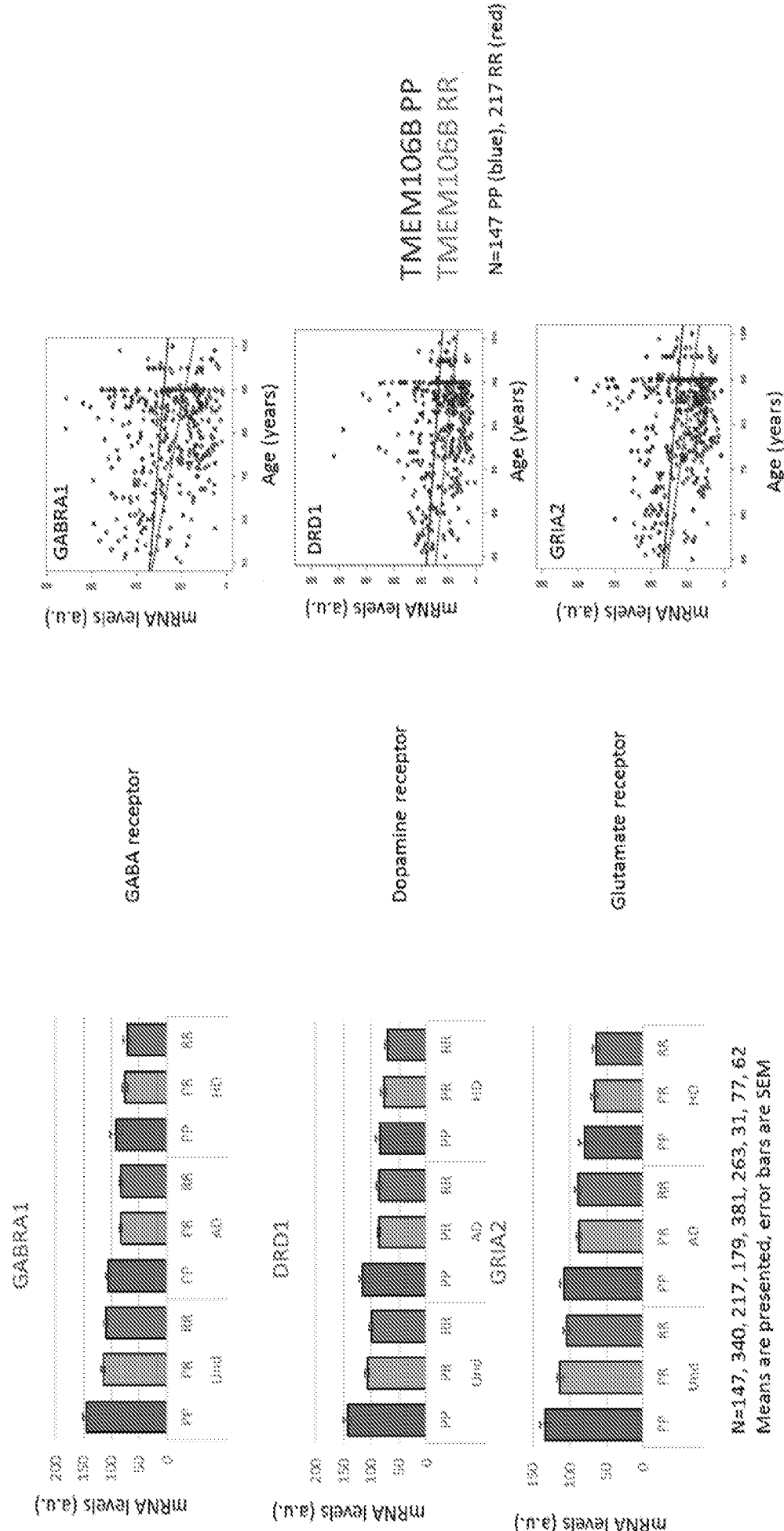
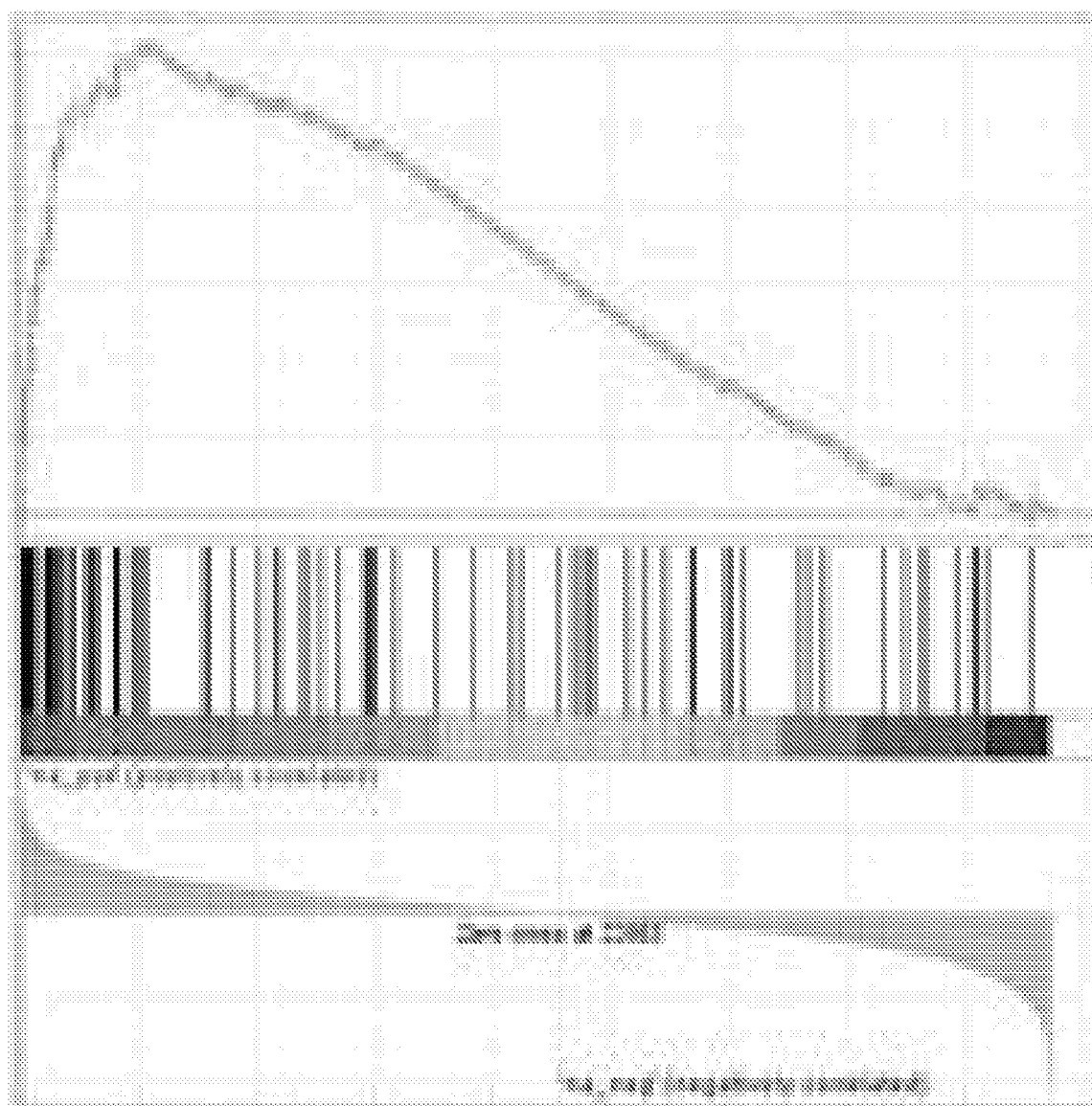


Fig. 50

**Fig. 51A**

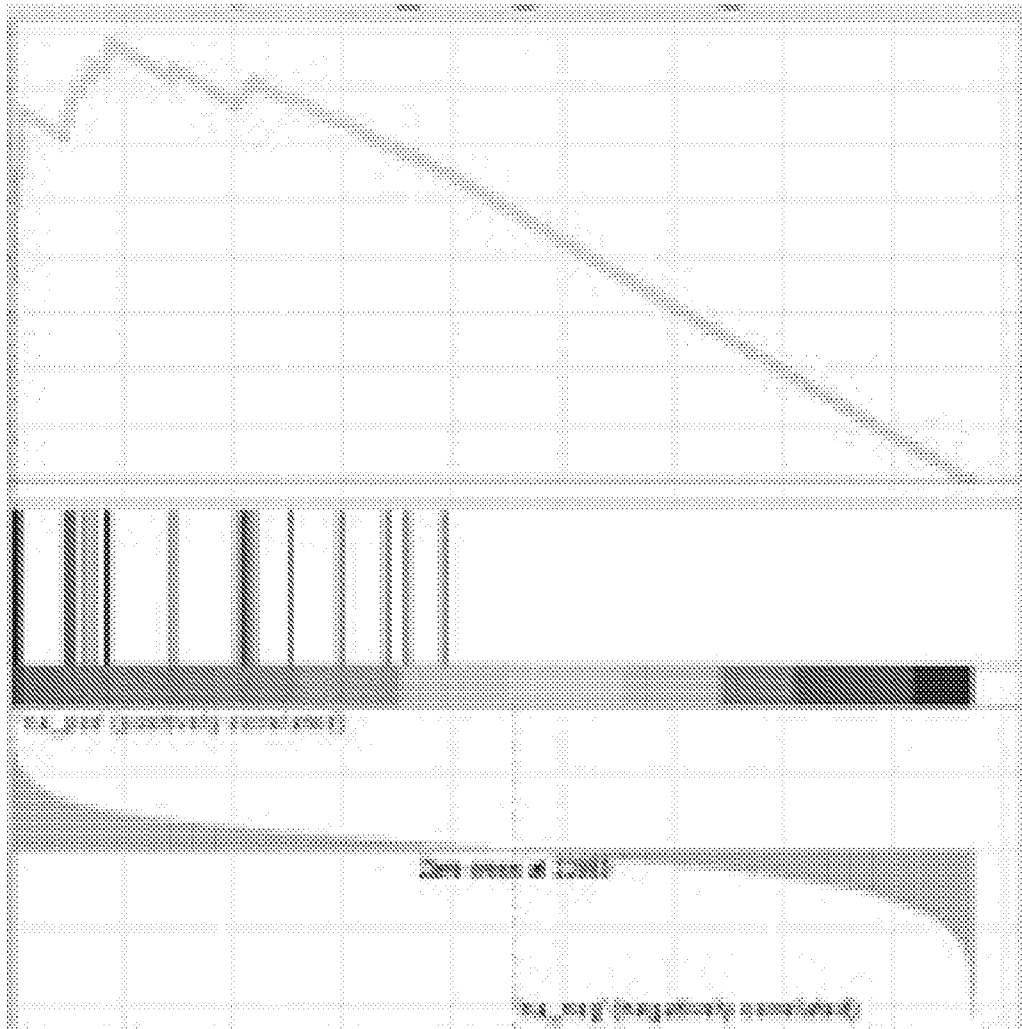
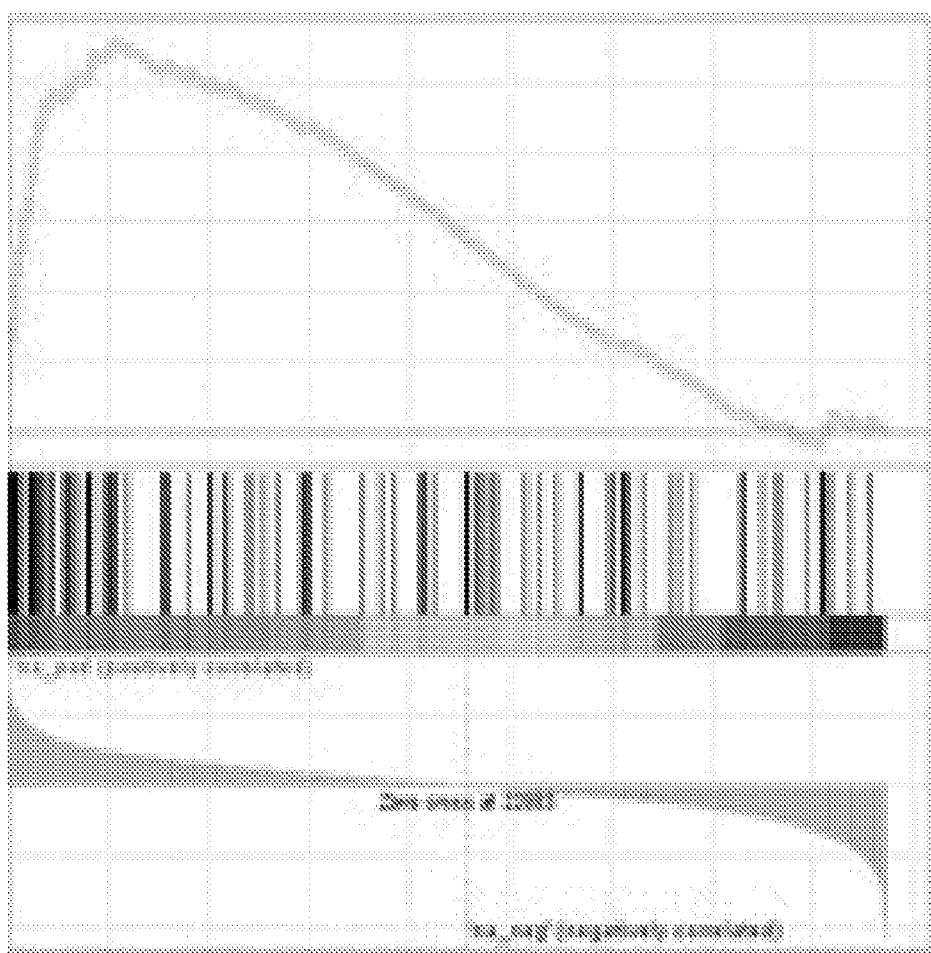
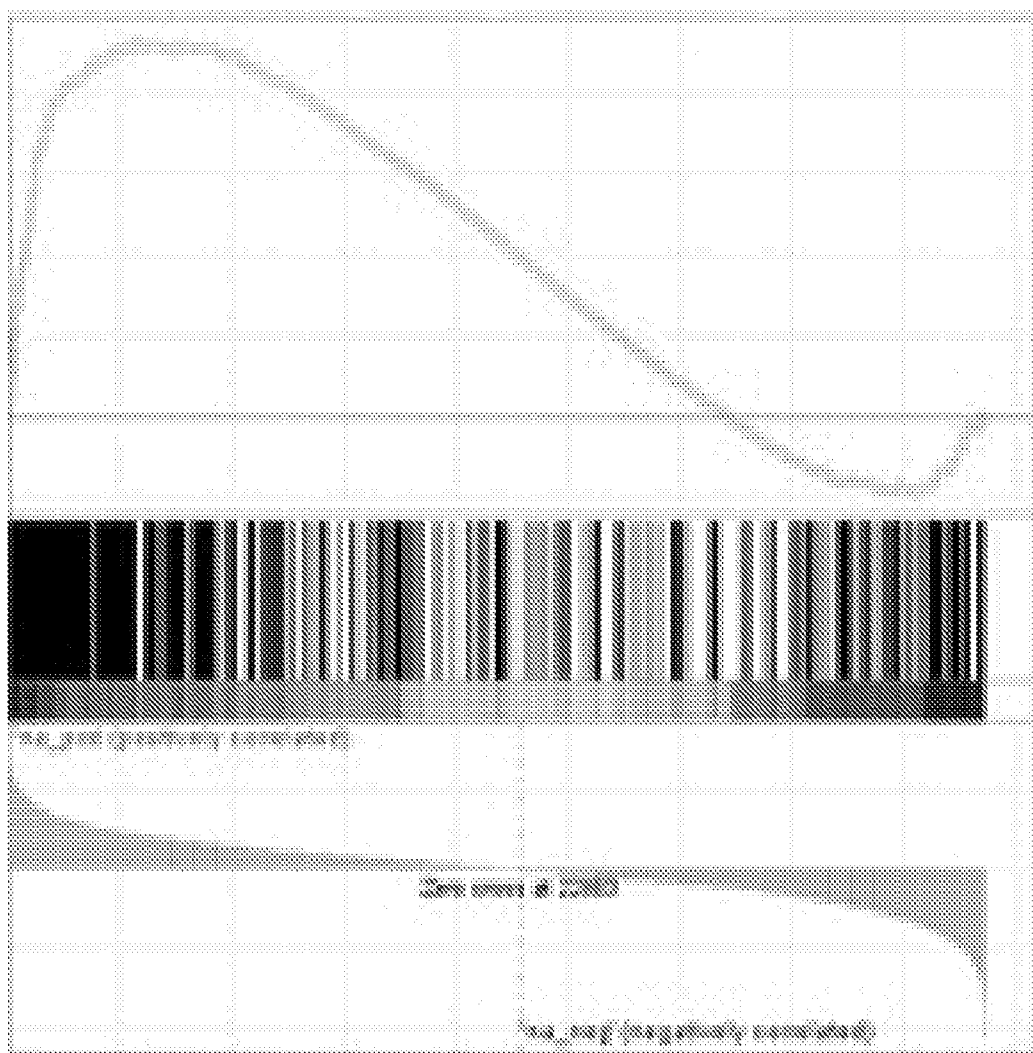
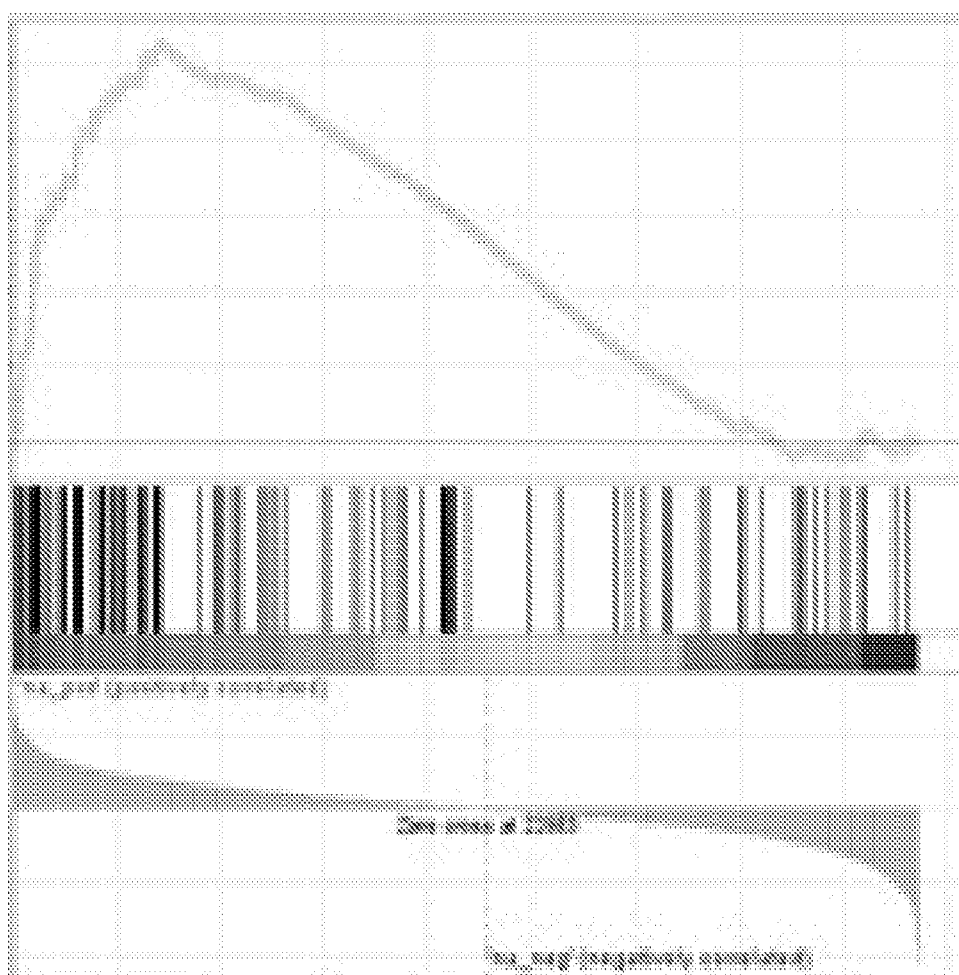
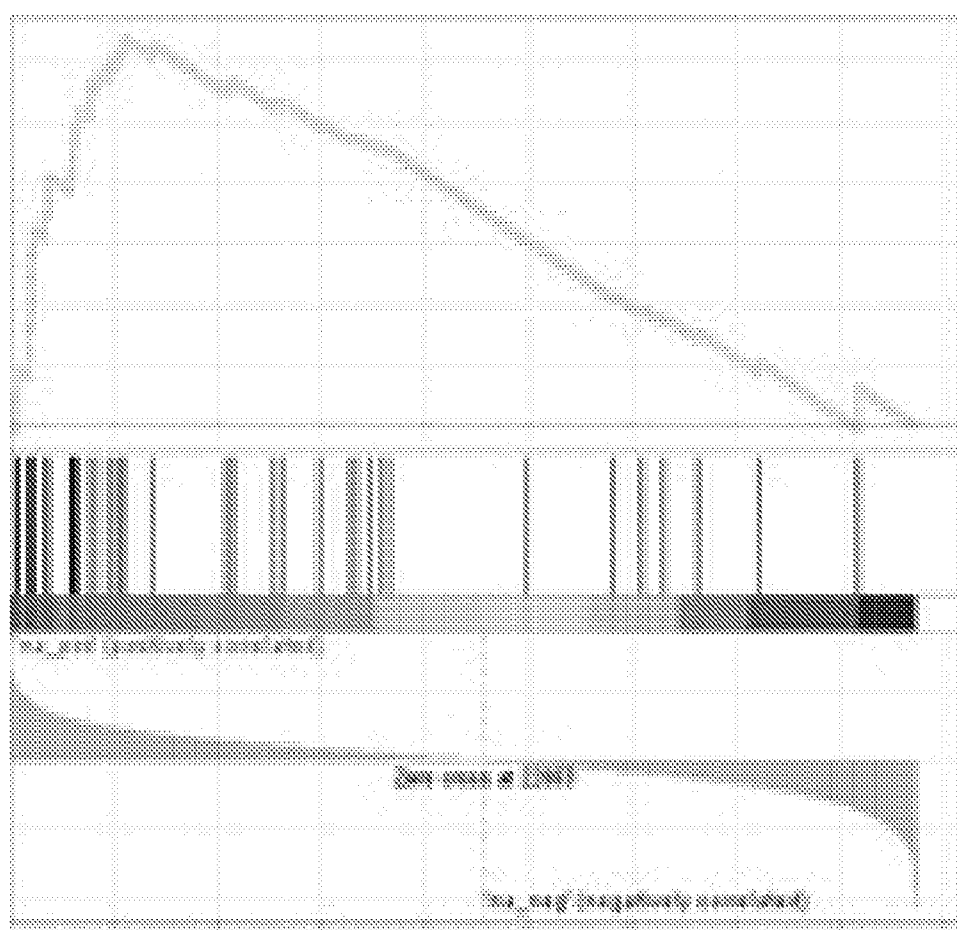


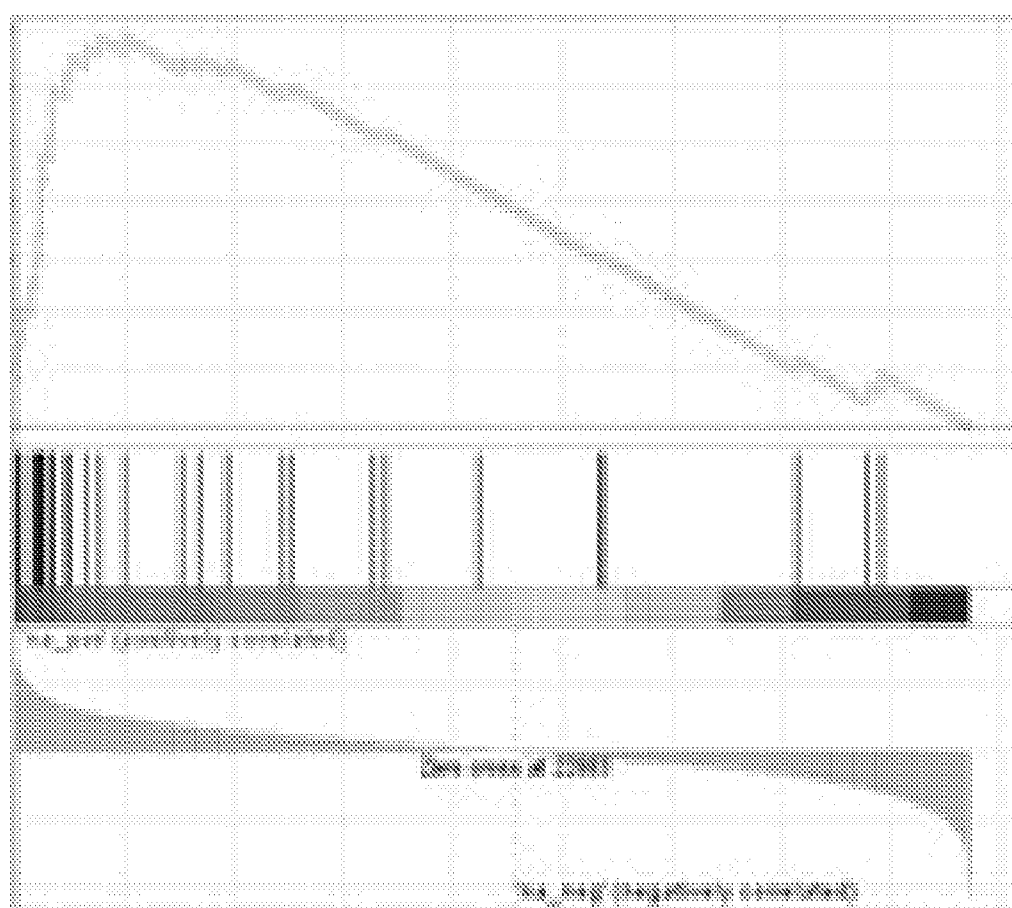
Fig. 51B

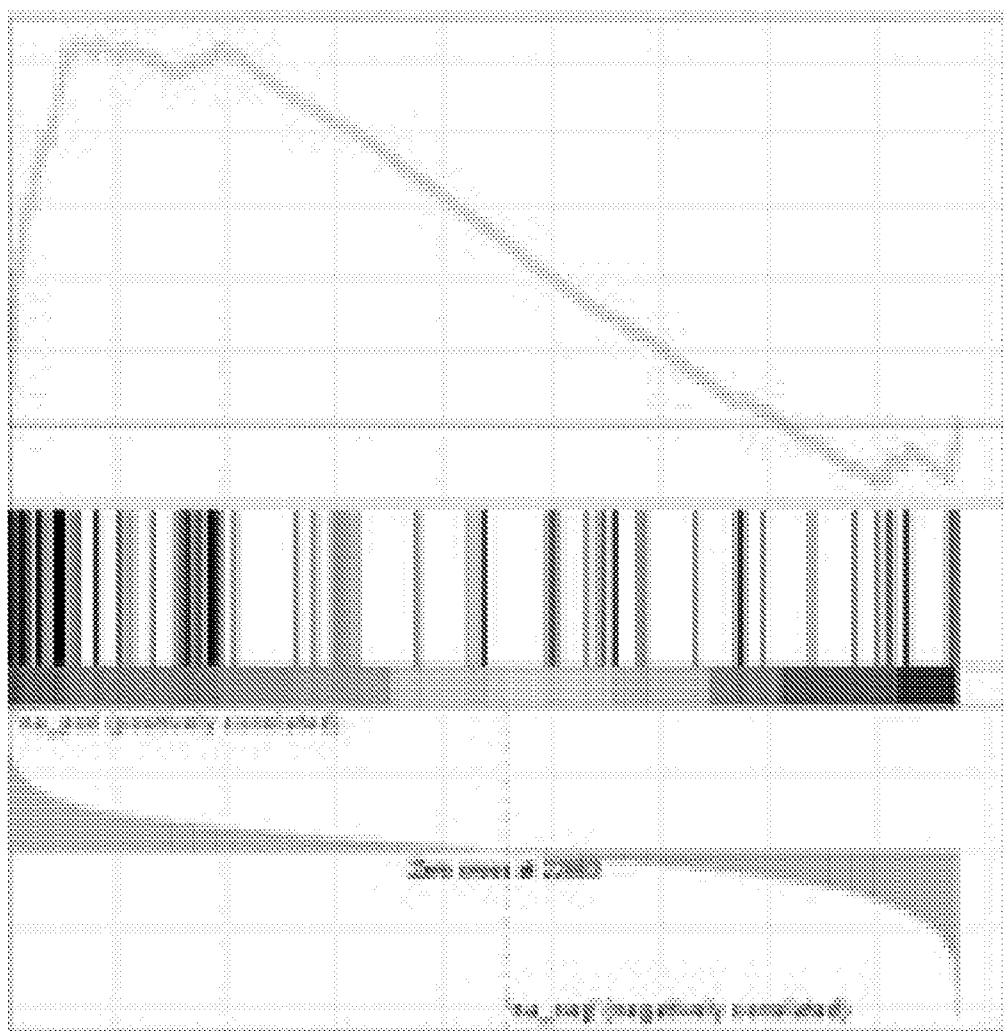
**Fig. 51C**

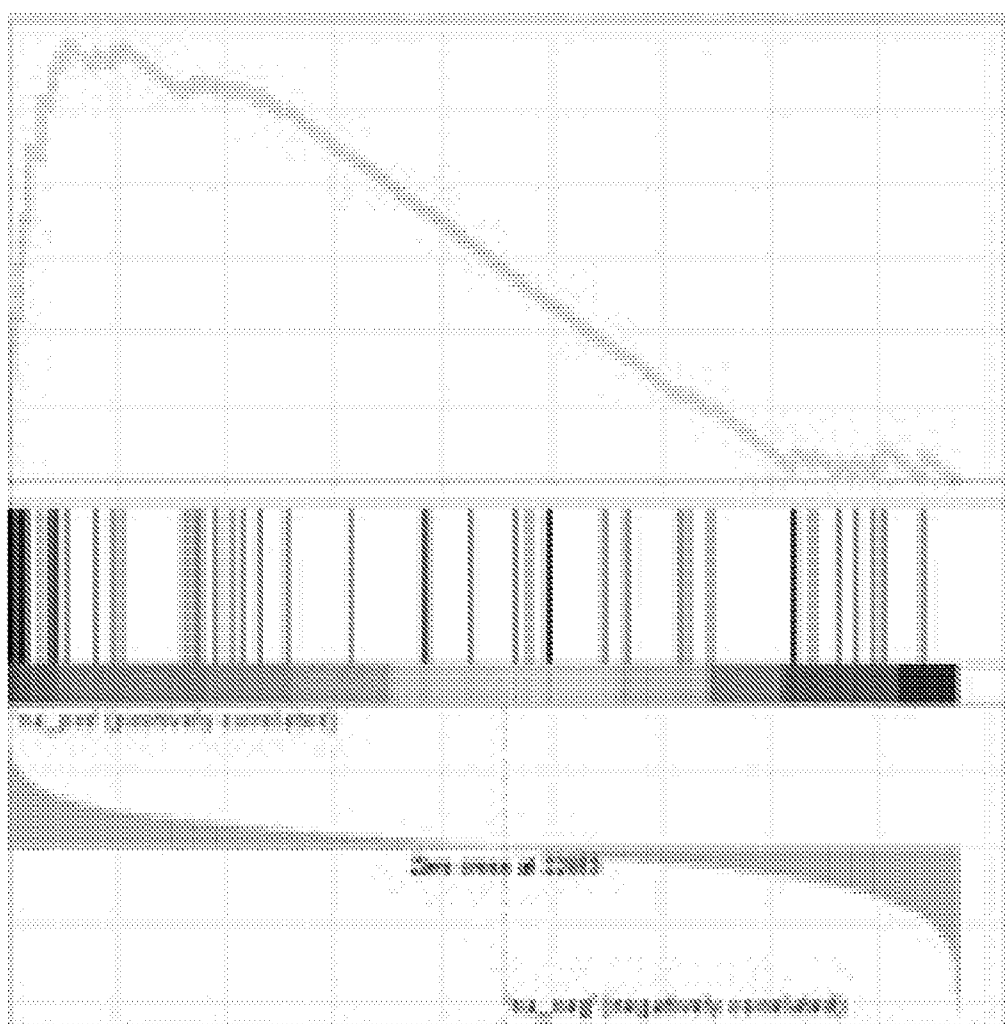
**Fig. 51D**

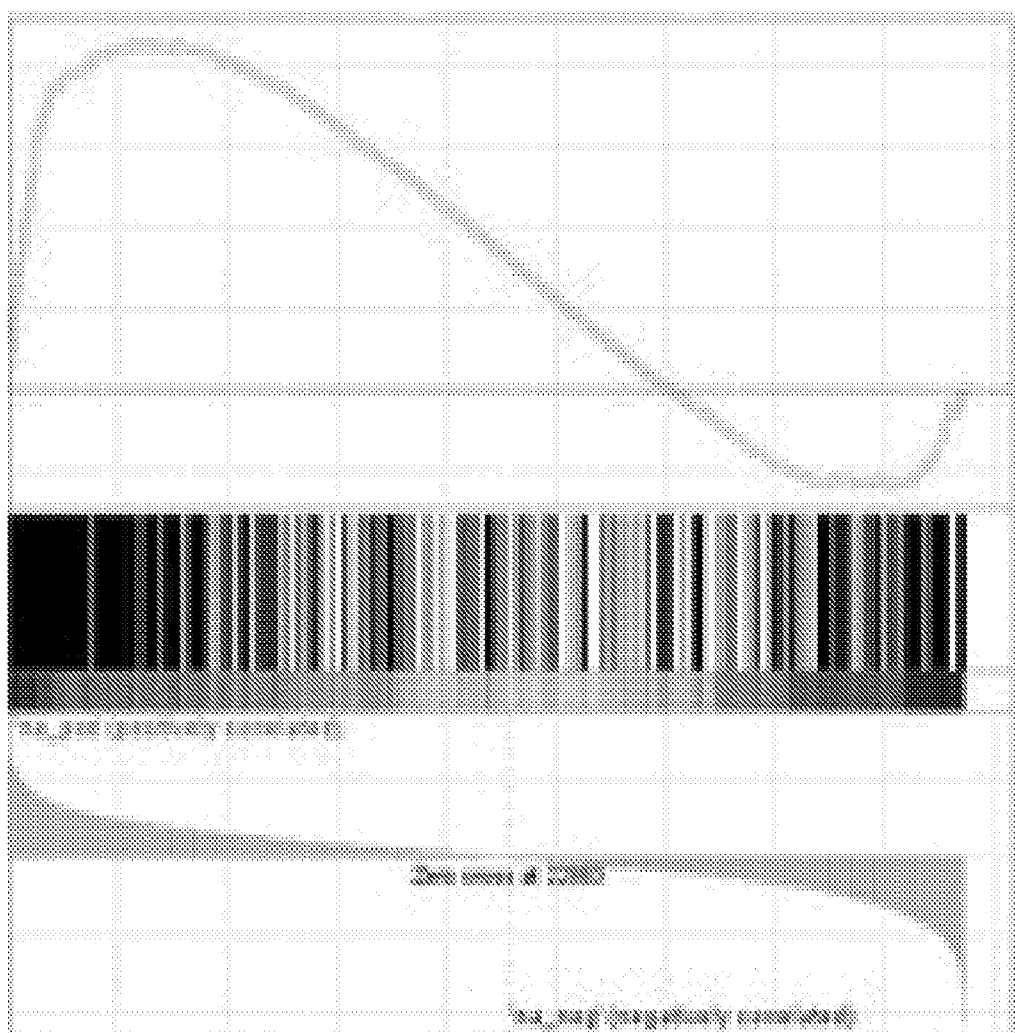
**Fig. 51E**

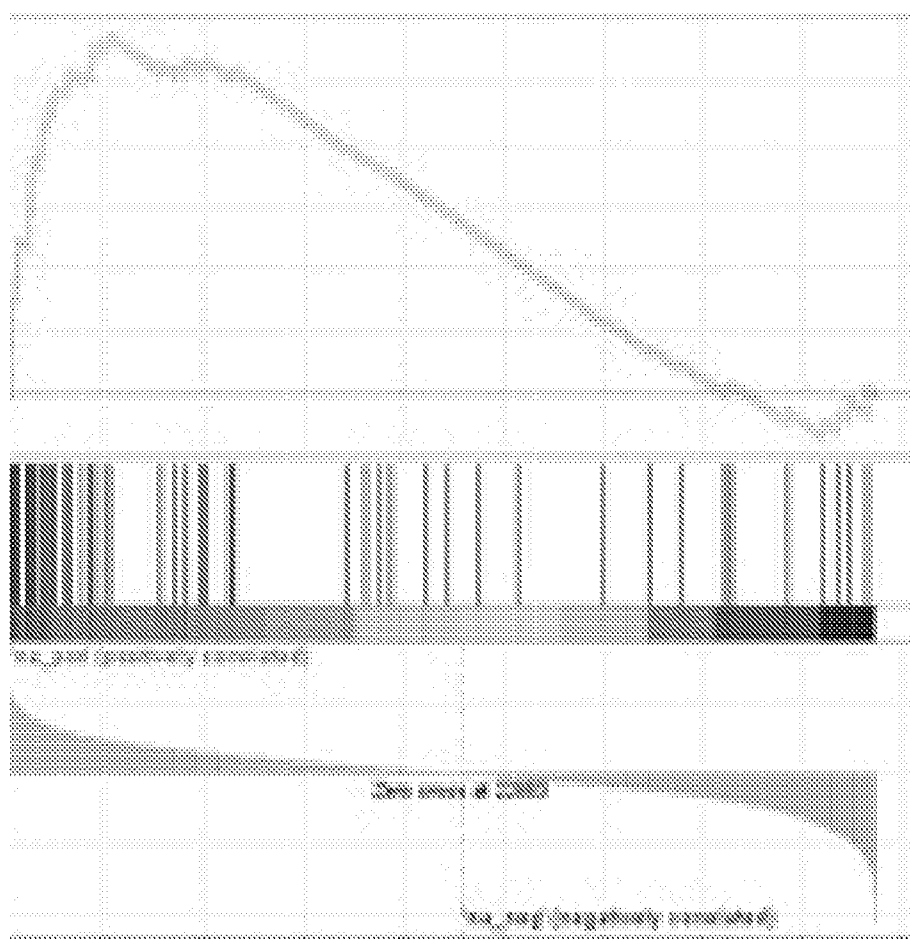
**Fig. 51F**

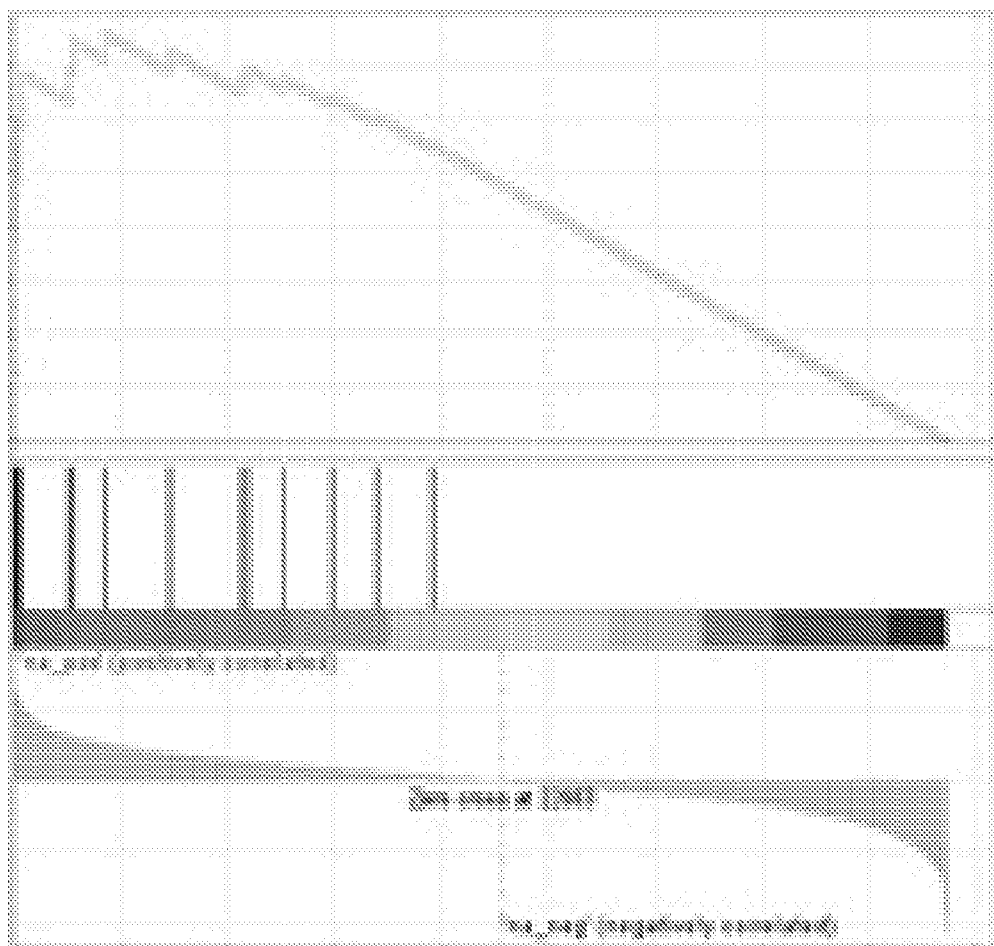
**Fig. 51G**

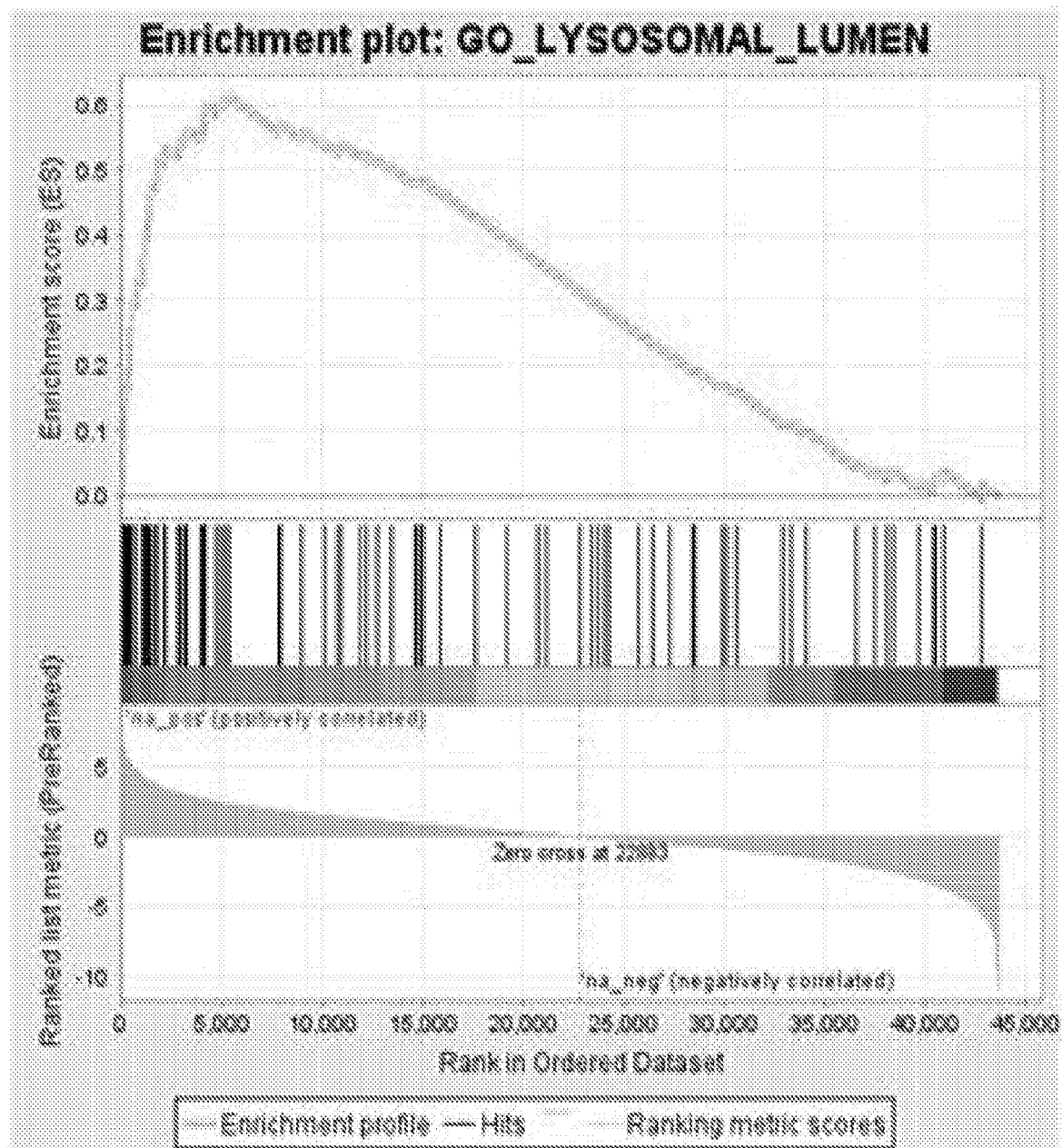
**Fig. 51H**

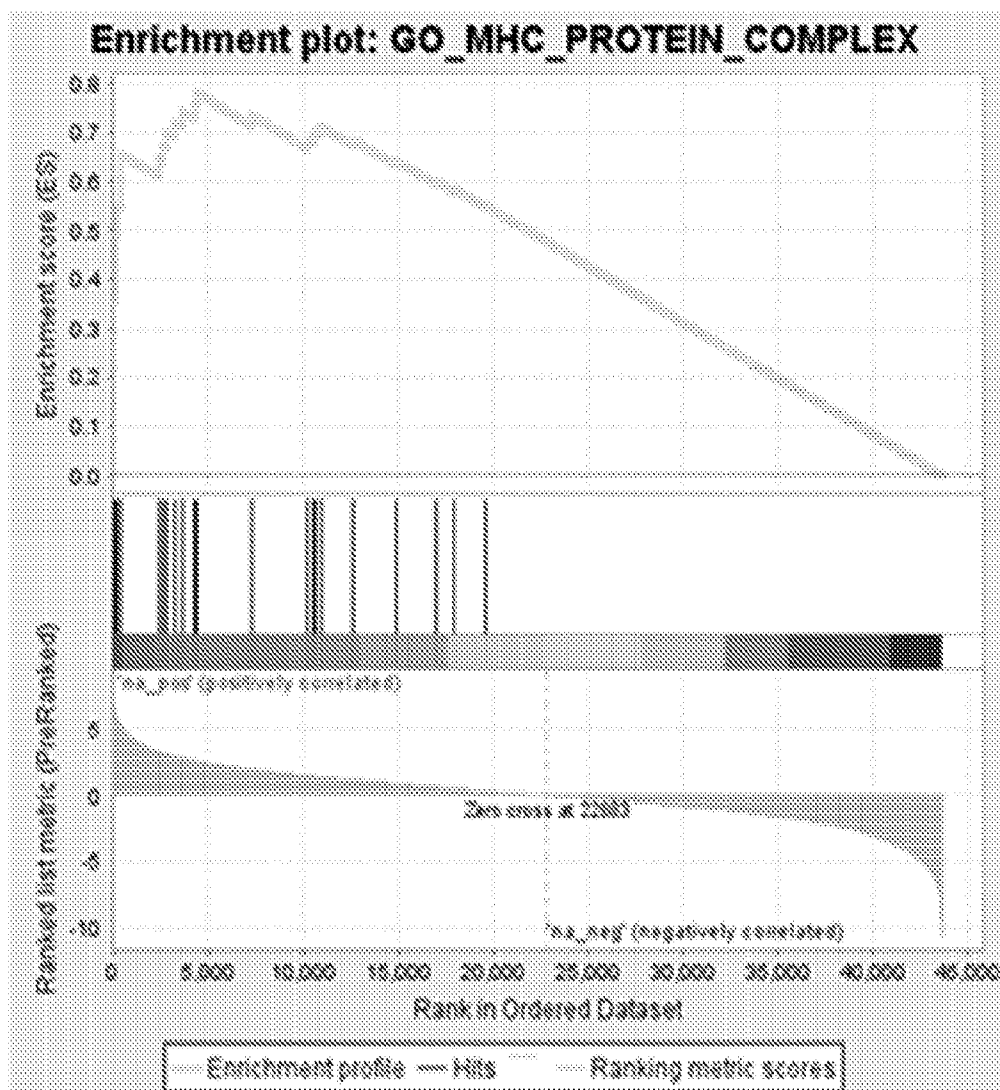
**Fig. 51I**

**Fig. 51J**

**Fig. 51K**

**Fig. 51L**

**Fig. 52**

**Fig. 52 cont.**

	PROB	CONF. NUMBER	CONF. TITLE	NAK. INQUIRY LIST	NAK. MEAS. SCORE	REMARKS	CONF. DETERMINED
1	GM2A			2	3.161	0.0353	Yes
2	CD74			40	6.995	0.0648	Yes
3	CSPG4			52	6.813	0.0941	Yes
4	EMASEI2			163	6.148	0.1182	Yes
5	ICN2			187	6.055	0.1439	Yes
6	GLUB			224	5.925	0.1688	Yes
7	GPC2			388	5.391	0.1884	Yes
8	MAN2B1			415	5.332	0.2110	Yes
9	LAMP2			483	5.210	0.2320	Yes
10	CISS			506	5.171	0.2539	Yes
11	PDGFRB			538	5.111	0.2754	Yes
12	IPP1			613	4.949	0.2951	Yes
13	GAA			731	4.744	0.3130	Yes
14	HSPG2			821	4.614	0.3310	Yes
15	VCAN			1036	4.321	0.3448	Yes
16	NAGLU			1107	4.221	0.3615	Yes
17	EGN			1140	4.184	0.3789	Yes
18	SDC2			1232	4.103	0.3946	Yes
19	IDUA			1286	4.051	0.4109	Yes
20	GALC			1344	3.985	0.4269	Yes

Fig. 53

21	SDC3			1374	3 952	0.4434	Yes
22	BCAN			1415	3 913	0.4534	Yes
23	GPC5			1507	3 842	0.4740	Yes
24	ARSA			1700	3 692	0.4856	Yes
25	SDC4			1828	3 595	0.4982	Yes
26	GYG2			1839	3 587	0.5136	Yes
27	NCAN			2138	3 417	0.5215	Yes
28	NEU4			2254	3 352	0.5334	Yes
29	ASAH1			2816	3 090	0.5339	Yes
30	HYAL1			2923	3 050	0.5447	Yes
31	GBA			3148	2 964	0.5524	Yes
32	SGSH			3264	2 925	0.5625	Yes
33	SCARB2			3867	2 693	0.5880	Yes
34	GYG1			4058	2 671	0.5675	Yes
35	GPC4			4070	2 667	0.5788	Yes
36	IF30			4179	2 634	0.5878	Yes
37	LUM			4192	2 630	0.5989	Yes
38	OMD			4746	2 481	0.5970	Yes
39	DCN			4899	2 442	0.6041	Yes
40	PRELP			5148	2 386	0.6087	Yes
41	TMDC5			5370	2 332	0.6137	Yes

Fig. 53 cont.

PROBE	GENE SYMBOL	GENE NAME	TRAINING GENE DIS	TRAINING SNP SCORE	ADJUDICATING ES	CORE ENRICHMENT
1 HLA-E			25	7.157	0.0750	Yes
2 CD74			40	6.995	0.1484	Yes
3 HLA-DRA			54	6.764	0.2195	Yes
4 HLA-DMA			70	6.621	0.2891	Yes
5 HLA-B			80	6.577	0.3582	Yes
6 HLA-F			101	6.454	0.4259	Yes
7 HLA-DMB			135	6.291	0.4915	Yes
8 HLA-DPA1			269	5.753	0.5492	Yes
9 B2M			416	5.330	0.6021	Yes
10 HLA-DPB1			489	5.199	0.6553	Yes
11 HLA-A			2401	3.295	0.6462	Yes
12 HLA-DOA			2652	3.160	0.6738	Yes
13 HLA-DRB1			2866	3.071	0.7014	Yes
14 HLA-C			3250	2.930	0.7235	Yes
15 MR1			3648	2.790	0.7438	Yes
16 HLA-DRE5			4292	2.599	0.7565	Yes
17 AZGP1			4425	2.566	0.7806	Yes

Fig. 53 cont.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 18/22678

A. CLASSIFICATION OF SUBJECT MATTER
IPC(8) - C12Q 1/68; G01N 33/50 (2018.01)
CPC - C12Q 2600/158, 1/6883, 1/6876; G01N 33/50

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 data base consulted during the international search (name of data base 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.
Y	Yang et al. Synchronized age-related gene expression changes across multiple tissues in human and the link to complex diseases. Sci Rep 19 October 2015 Vol 5 No 15145 Pages 1-16. Especially pg 4 fig 2 A,B; pg 6 para 2; pg 10 para 7.	1-3, 5, 6, 13-15, 17-18
Y	WO 2001/020043 A1 (Affymetrix, Inc.) 22 March 2001 (22.03.2001). Especially pg 14 ln 22-32.	1-3, 5, 6, 13-15, 17-18
A	Zahn et al. AGEMAP: a gene expression database for aging in mice. PLoS Genet November 2007 Vol 3 No 11 Pages e201 1-12. Especially abstract.	1-3, 5, 6, 13-15, 17-18
A	de Magalhaes et al. Meta-analysis of age-related gene expression profiles identifies common signatures of aging. Bioinformatics 01 April 2009 Vol 25 No 7 Pages 875-881. Especially abstract.	1-3, 5, 6, 13-15, 17-18

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

☐ See patent family annex.

* Special categories of cited documents:

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

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

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

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

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

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

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

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

"&" document member of the same patent family

Date of the actual completion of the international search

20 June 2018

Date of mailing of the international search report

26 JUL 2018

Name and mailing address of the ISA/US

Mail Stop PCT, Attn: ISA/US, Commissioner for Patents
P.O. Box 1450, Alexandria, Virginia 22313-1450
Facsimile No. 571-273-8300

Authorized officer:

Lee W. Young

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

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 18/22678

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.: 4,7-12,16,19-24,28-36,40-48, 54-65,71-82,86-103
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:
-----Go to Extra Sheet for continuation-----

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.:
Claims 1-3, 5, 6, 13-15, 17-18

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.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 18/22678

continuation of Box III: Observations where Unity of Invention is lacking

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

Group I: Claims 1-3, 5, 6, 13-15, 17-18, drawn to a method of determining a biological age of a sample from a subject.

Group II: Claims 25-27, 37-39, drawn to a method of determining a phenotype of a sample from a subject, wherein the phenotype is correlated with a haplotype of interest.

Group III: Claims 49-53, 66-70, drawn to a method of identifying one or more genetic modifiers of a plurality of genes whose expression level is correlated with a genotype of interest.

Group IV: Claims 83-85, 104-106, drawn to a method of administering one or more genetic modifiers to a subject.

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

Special Technical Features:

Group I has the special technical feature of determining for each gene whose expression level is significantly correlated with chronological age, a residual value, not required by Groups II-IV.

Group I has the special technical feature of performing a linear regression of the expression level of each gene of the plurality of genes as a function of the chronological age of each gene of the plurality of genes to determine a linear regression coefficient for each gene, not required by Groups II-IV.

Group I has the special technical feature where the chronological age of an individual or plurality of individuals is known, not required by Groups II-IV.

Group I has the special technical feature of an integrated ratio that corresponds to the biological age of the sample, not required by Groups II-V.

Group II has the special technical feature of determining, for each gene whose expression level is significantly correlated with the haplotype of interest, a residual value, not required by Groups I, III, IV.

Group II has the special technical feature of performing a linear regression of the expression level of each gene of the plurality of genes as a function of the haplotype of interest for each of the plurality of individuals to determine a linear regression coefficient for each gene, not required by Groups I, III, IV.

Group II has the special technical feature of providing a gene expression level of a plurality of genes in a sample for a plurality of individuals, wherein each individual's haplotype of interest is known, not required by Groups I, III, IV.

Group II has the special technical feature of an integrated ratio that corresponds to the phenotype of the sample, not required by Groups I, III, IV.

Group III has the special technical feature of performing a linear regression of the expression level of each gene of the plurality of genes as a function of the genotype of interest for each of the plurality of individuals to determine a linear regression coefficient for each gene, not required by Groups I, III, IV.

Group III has the special technical feature of determining, for each gene whose expression level is significantly correlated with the genotype of interest, a residual value, not required by Groups I, II, IV.

Group III has the special technical feature of providing a gene expression level of a plurality of genes in a sample for a plurality of individuals, wherein each individual's genotype of interest is known, not required by Groups I, II, IV.

Group III has the special technical feature of an integrating the ratios of each gene that correlates to the genotype of interest, not required by Groups I, II, IV.

Group III has the special technical feature of identifying genetic modifiers that modulated the integrated ratio value, not required by Groups I, II, IV.

Group IV has the special technical feature of administering one or more genetic modifiers to a subject.

-----continued on next sheet-----

-----continued from previous sheet-----

Common Technical Features:

Groups I-III share the common technical features of:

1. A computer implemented method or computer program product comprising instructions which, when the program is executed by a computer, cause the computer to carry out specific steps.
2. Providing a gene expression level of a plurality of genes in a sample from a subject or plurality of subjects.
3. Determining, for each gene whose expression level is significantly correlated with a chosen parameter (e.g. chronological age, genotype, phenotype, etc.), a residual value, wherein the residual value is determined by comparing the gene expression level of said gene in the sample from the subject to a linear regression of the expression level of the gene as a function a chosen parameter, for a plurality of individuals.
4. Determining for each gene that is correlated with a chosen parameter, a ratio of the residual value to the regression coefficient [i.e. slope "m" of linear regression curve, $y=mx + b$] for the gene.

Groups III and IV share the common technical feature of:

5. a genetic modifier or modulator of a gene or plurality of genes.

However, said common technical features do not represent a contribution over the prior art, and are obvious over WO 2001/020043 A1 to Affymetrix, Inc. (hereinafter "Affymetrix"), in view of the publication titled "Synchronized age-related gene expression changes across multiple tissues in human and the link to complex diseases" by Yang et al. (hereinafter "Yang") [published 19 October 2015 in Sci Rep Vol 5 No 15145 Pages 1-16].

As to common technical feature #1, A computer implemented method or computer program product comprising instructions which, when the program is executed by a computer, cause the computer to carry out specific steps, Affymetrix teaches (pg 14 ln 22-32; "Some embodiments of the computer program product of the invention contains computer codes for displaying data using the principal component method of the invention. The computer program product comprises: a) computer code that receives a plurality of biological profiles [i.e. gene expression profiles], each of the profiles comprises a plurality of biological variables [i.e. individual genes]; b) computer code that reduces the number of the biological variables to obtain a set of reduced profiles; c) computer code that performs a principal component analysis of the reduced profiles; d) computer code that displays the biological variables according a first axis and a second axis, wherein the first axis is the first component of the principal component analysis and the second axis is the second component of the principal component analysis, and wherein the first axis is perpendicular to the second axis; and e) computer readable medium that stores the computer codes. IV. Gene Expression Monitoring Methods").

As to common technical feature #2, Providing a gene expression level of a plurality of genes in a sample from a subject or plurality of subjects, Affymetrix teaches (claim 1; "1. A method for analyzing a plurality of biological profiles comprising: a. providing said plurality of biological profiles, wherein each of said profiles comprises a plurality of biological variables; b. reducing the number of said biological variables to obtain a set of reduced profiles; and c. subjecting said reduced profiles to statistical analysis"; claim 15; "15. The method of Claim 1 wherein said biological profiles are gene expression profiles and each of said biological variables represents the expression of a gene").

Additionally, as to common technical feature #2, Yang teaches (pg 4 Fig. 2A; "Age-associated gene expression in subcutaneous adipose tissue: (a) Heatmap of 1,134 age associated genes (row) on 94 samples (column). Colors represent normalized gene expression values with blue for low expression and red for high expression. The age of each individual is displayed at the bottom and also illustrated in color bar at the top with dark green for young and yellow for old").

As to common technical feature #3, Determining, for each gene whose expression level is significantly correlated with a chosen parameter (e.g. chronological age, genotype, phenotype, etc.), a residual value, wherein the residual value is determined by comparing the gene expression level of said gene in the sample (pg 6 para 2; "As shown in Fig. 2a and Supplementary Fig. S1, although the chronological ages of the tissue donors are largely consistent with the "young" or "old" clustering, there are some apparent "outliers". For example, the 15th adipose sample (from the left) in Fig. 2a is from an individual with an old chronological age of 70 but is clustered with the "young" group where most donors range in age from 20 to 50. This indicated that the molecular age of adipose tissue from this individual is about 30 years younger than its chronological age") from the subject to a linear regression of the expression level of the gene as a function a chosen parameter, for a plurality of individuals (pg 10 para 7; "Linear regression model for aging gene detection. In each tissue, we modelled gene expression using the following linear regression model"; pg 4 Fig. 2B; b) Scatter plot of 2 representative age-associated gene expression patterns PYH1N1 and EIF5AL1 in adipose tissue. Pearson-R value in the title represents the Pearson correlation coefficient between gene expression and age across all samples. The solid blue triangles plot male samples and solid red circle female samples. Similarly, the blue and red lines denote the REGRESSION lines for male and female samples, respectively").

As to common technical feature #4, concerning determining for each gene that is correlated with a chosen parameter, a ratio of the residual value to the regression coefficient [i.e. slope "m" of linear regression curve, $y=mx + b$] for the gene, based on the teaching of Yang as presented, for common technical feature #3, it would have been obvious that from the residual value of "outlier" subject 15 in #3 above, the X-axis apparent age (as opposed to chronological age) of the subject tissue gene could have been calculated or determined from where particular residual value as plotted on the X-axis where it intersected on the regression line of Fig. 2B.

-----continued on next sheet-----

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 18/22678

-----continued from previous sheet-----

As to common technical feature #5, a genetic modifier or modulator of a gene or plurality of genes, Affymetrix teaches (pg 4 ln 24-29; "In an additional aspect of the invention, methods for studying a plurality of drugs are provided. In some embodiments, the expression of more than 50 genes, preferably more than 1000 genes, more preferably more than 2000 genes and most preferably more than 4000 genes, in a biological sample in response to a plurality of drugs is measured, preferably by nucleic acid probe arrays, to obtain a plurality of gene expression profiles, each of the profiles representing the response of the biological sample to one of the plurality of drugs").

As the common technical features were known in the art at the time of the invention, they cannot be considered common special technical features that would otherwise unify the groups. The inventions lack unity with one another.

Therefore, Groups I-IV lack unity of invention under PCT Rule 13 because they do not share a same or corresponding special technical feature.

Note concerning claim 105: Claim 105 is written to depend from claim 86, but more appropriately should depend from claim 83, because claim 83 has antecedent precedent. For the purposes of the International Search & Opinion, claim 105 is interpreted to depend from claim 83.

Item 4 (contd.) : 6.4

Claims 4, 7-12, 16, 19-24, 28-36, 40-48, 54-65, 71-82, 86-103 are held unsearchable because they are not drafted according to the second and third sentences of PCT Rule 6.4(a).