



(12)发明专利申请

(10)申请公布号 CN 105745333 A

(43)申请公布日 2016.07.06

(21)申请号 201380070066.1

史蒂芬·H·弗兰德

(22)申请日 2013.11.12

(74)专利代理机构 北京中恒高博知识产权代理有限公司 11249

(30)优先权数据

代理人 夏晏平

61/724,528 2012.11.09 US

(85)PCT国际申请进入国家阶段日

(51)Int.Cl.

2015.07.09

C12Q 1/68(2006.01)

(86)PCT国际申请的申请数据

G12N 15/11(2006.01)

PCT/US2013/069710 2013.11.12

G01N 33/48(2006.01)

(87)PCT国际申请的公布数据

G01N 33/53(2006.01)

W02014/075083 EN 2014.05.15

(71)申请人 加利福尼亚大学董事会

地址 美国加利福利亚州

申请人 赛智生物网络

(72)发明人 康·张 格里高利·汉恩

崔伊·伊德克 贾斯廷·吉尼

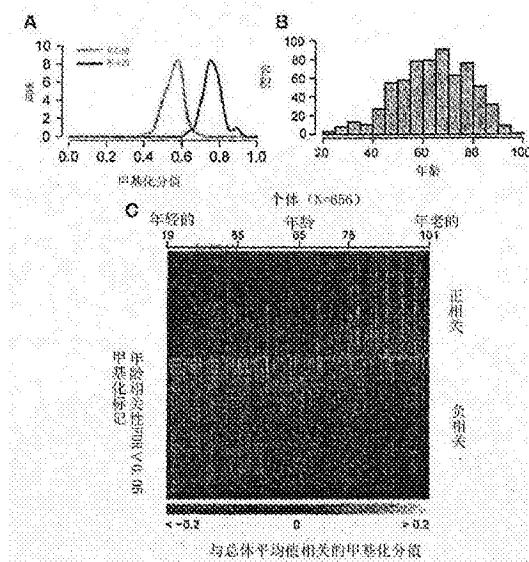
权利要求书10页 说明书35页 附图49页

(54)发明名称

用于预测年龄和鉴别诱发或者抑制早衰试剂的方法

(57)摘要

本发明提供了基于受试者的表观基因组预测受试者年龄的方法。



1. 一种基于受试者表观基因组预测受试者年龄的方法,该方法包括:

(a)获得受试者的生物样品;

(b)确定如图9、表S3、S4和/或S5任一所示的在受试者表观基因组中与年龄相关的一套表观遗传标记的甲基化状态;及

(c)对比受试者与年龄相关的一套表观遗传标记的甲基化状态和年龄相关的参考人群的相同标记的甲基化状态,以获得用于预测受试者年龄的一个值或者一个值的范围,从而基于受试者的表观基因组预测受试者的年龄。

2. 权利要求1的方法,其中在(b)中,确定甲基化状态包括:

(a)从样品中分离基因组DNA或核DNA;

(b)使分离的基因组DNA或核DNA与一种或多种的探针反应,所述探针与未修饰的胞嘧啶的做出不同的反应,以便胞嘧啶转化为尿嘧啶;

(c)通过检测胞嘧啶或尿嘧啶的存在,分析在样品分离的基因组DNA或核DNA中的CpG二核苷酸胞嘧啶位置上的甲基化状态,胞嘧啶或尿嘧啶的存在表明在原CpG二核苷酸中分别存在5-甲基胞嘧啶或未经修饰的胞嘧啶;及

(d)从(c)中确定5-甲基胞嘧啶或未经修饰的胞嘧啶最初呈现在各自与年龄相关的表观遗传标记的比例,或者,可选择地,从(c)中确定最初呈现在各自与年龄相关的表观遗传标记5-甲基胞嘧啶与未经修饰的胞嘧啶的比或者未经修饰的胞嘧啶与5-甲基胞嘧啶的比。

3. 权利要求1的方法,其中在(b)中,确定甲基化状态包括:

(a)从样品中分离基因组DNA或核DNA;

(b)用一种或多种限制性内切酶培育分离的基因组DNA或核DNA,所述的限制性内切酶(i)可识别特定DNA序列;(ii)受到CpG二核苷酸的影响,在内部或接近限制性内切酶的鉴别点或者分裂位点,及(iii)基于在所述CpG二核苷酸的胞嘧啶的C-5位位置上存在或不存在甲基基团DNA裂解方式的不同;

(c)在样品中分离的基因组DNA或核DNA中,通过表明在内部或接近限制性内切酶的鉴别点或者分裂位点的原CpG二核苷酸中存在5-甲基胞嘧啶的限制性内切酶在潜在的分裂位点的抗分裂,分析CpG二核苷酸C位的甲基化状态,限制性内切酶的分裂敏感性可以表明存在未修饰的胞嘧啶;

(d)从(c)中确定5-甲基胞嘧啶或未经修饰的胞嘧啶最初呈现在各自与年龄相关的表观遗传标记的比例,或者,可选择地,从(c)中确定最初呈现在各自与年龄相关的表观遗传标记5-甲基胞嘧啶与未经修饰的胞嘧啶的比或者未经修饰的胞嘧啶与5-甲基胞嘧啶的比。

4. 权利要求1的方法,在(b)中,由选自于图9,表格S3、S4和/或S5受试者表观基因组的5个或以上的与年龄相关的表观遗传标记;选自图9,表格S3、S4和/或S5受试者表观基因组的10个或以上的与年龄相关的表观遗传标记;选自图9,表格S3、S4和/或S5受试者表观基因组的15个或以上的与年龄相关的表观遗传标记;选自图9,表格S3、S4和/或S5受试者表观基因组的20个或以上的与年龄相关的表观遗传标记;选自图9,表格S3、S4和/或S5受试者表观基因组的25个或以上的与年龄相关的表观遗传标记;选自图9,表格S3、S4和/或S5受试者表观基因组的30个或以上的与年龄相关的表观遗传标记;选自图9,表格S3、S4和/或S5受试者表观基因组的35个或以上的与年龄相关的表观遗传标记;选自图9,表格S3、S4和/或S5受试者表观基因组的40个或以上的与年龄相关的表观遗传标记;选自图9,表格S3、S4和/或S5受试

者表观基因组的45个或以上的与年龄相关的表观遗传标记;选自图9,表格S3、S4和/或S5受试者表观基因组的50个或以上的与年龄相关的表观遗传标记;选自图9,表格S3、S4和/或S5受试者表观基因组的55个或以上的与年龄相关的表观遗传标记;选自图9,表格S3、S4和/或S5受试者表观基因组的60个或以上的与年龄相关的表观遗传标记;选自图9,表格S3、S4和/或S5受试者表观基因组的65个或以上的与年龄相关的表观遗传标记;选自图9,表格S3、S4和/或S5受试者表观基因组的70个或以上的与年龄相关的表观遗传标记,可确定甲基化状态。

5. 权利要求1的方法,在(b)中,基于图9或表S3中受试者表观基因组的5个或5个以上的与年龄相关的表观遗传标记;图9或表S3受试者表观基因组的10个或10个以上的与年龄相关的表观遗传标记;图9或表S3受试者表观基因组的15个或15个以上的与年龄相关的表观遗传标记;图9或表S3受试者表观基因组的20个或20个以上的与年龄相关的表观遗传标记;图9或表S3受试者表观基因组的25个或25个以上的与年龄相关的表观遗传标记;图9或表S3受试者表观基因组的30个或30个以上的与年龄相关的表观遗传标记;图9或表S3受试者表观基因组的35个或35个以上的与年龄相关的表观遗传标记;图9或表S3受试者表观基因组的40个或40个以上的与年龄相关的表观遗传标记;图9或表S3受试者表观基因组的45个或45个以上的与年龄相关的表观遗传标记;图9或表S3受试者表观基因组的50个或50个以上的与年龄相关的表观遗传标记;图9或表S3受试者表观基因组的55个或55个以上的与年龄相关的表观遗传标记;图9或表S3受试者表观基因组的60个或60个以上的与年龄相关的表观遗传标记;图9或表S3受试者表观基因组的65个或65个以上的与年龄相关的表观遗传标记,或者图9或表S3受试者表观基因组的70个或70个以上的与年龄相关的表观遗传标记,可确定甲基化状态。

6. 权利要求1的方法,其中受试者的年龄为实足年龄或者生物年龄。

7. 权利要求1的方法,其中,表观基因组是脱氧核糖核酸(DNA),所述的DNA经受表观遗传修饰。

8. 权利要求1的方法,其中,表观遗传修饰为CpG残留物的甲基化。

9. 权利要求1或8的方法,其中,甲基化是甲基基团共价键合在胞嘧啶碳-5位置上。

10. 权利要求1的方法,其中,生物样本为血液、淋巴细胞、单核细胞、中性粒细胞、嗜碱性粒细胞、嗜酸性粒细胞、髓系细胞、淋巴系细胞、骨髓、唾液、腮拭子、鼻拭子、尿、粪便、毛、乳腺组织、卵巢组织、子宫组织、宫颈组织、前列腺组织、睾丸组织、脑组织、神经细胞、星形胶质细胞、肝组织、肾、甲状腺组织、胃组织、肠组织、胰腺组织、血管组织、皮肤、肺组织、骨组织、软骨、韧带、肌腱、脂肪细胞、肌细胞、神经元、星形胶质细胞、具有不同传代次数的培养细胞、癌症/肿瘤细胞、癌症/肿瘤组织、正常细胞、正常组织、含有遗传物质核的任一组织或细胞或DNA的形式的遗传物质中的任何一种。

11. 权利要求1的方法,其中,与年龄相关的表观遗传标记包括一个CpG残留物,其中,胞嘧啶C-5位的甲基化随受试者相关的物种的实足年龄而变化。

12. 权利要求1的方法,其中,与受试者相关的物种是人类。

13. 权利要求1的方法,其中,所述与年龄相关的一套表观遗传标记包括在受试者基因组中受胞嘧啶C-5位年龄依赖性甲基化的影响的个体CpG残留物。

14. 权利要求13的方法,其中,所述的一套标记包括约70个含有与年龄相关的表观遗传

标记独特的CpG残留物。

15. 权利要求13的方法,其中,所述的一套标记包括一个或多个如图9、表S3、表S4或表S5所示的标记。

16. 权利要求15的方法,其中,所述的一套标记包括如图9或表S3所示的5个或5个以上的与年龄相关的表观遗传标记;如图9或表S3所示的10个或10个以上的与年龄相关的表观遗传标记;如图9或表S3所示的15个或15个以上的与年龄相关的表观遗传标记;如图9或表S3所示的20个或20个以上的与年龄相关的表观遗传标记;如图9或表S3所示的25个或25个以上的与年龄相关的表观遗传标记;如图9或表S3所示的30个或30个以上的与年龄相关的表观遗传标记;如图9或表S3所示的35个或35个以上的与年龄相关的表观遗传标记;如图9或表S3所示的40个或40个以上的与年龄相关的表观遗传标记;如图9或表S3所示的45个或45个以上的与年龄相关的表观遗传标记;如图9或表S3所示的50个或50个以上的与年龄相关的表观遗传标记;如图9或表S3所示的55个或55个以上的与年龄相关的表观遗传标记;如图9或表S3所示的60个或60个以上的与年龄相关的表观遗传标记;如图9或表S3所示的65个或65个以上的与年龄相关的表观遗传标记或者如图9或表S3所示的70个或70个以上的与年龄相关的表观遗传标记。

17. 权利要求15的方法,其中,所述的一组标记包括如表格S5所示的5个或5个以上的与年龄相关的表观遗传标记;如表S5所示的10个或10个以上的与年龄相关的表观遗传标记;如表S5所示的15个或15个以上的与年龄相关的表观遗传标记;如表S5所示的20个或20个以上的与年龄相关的表观遗传标记;如表S5所示的25个或25个以上的与年龄相关的表观遗传标记;如表S5所示的30个或30个以上的与年龄相关的表观遗传标记;如表S5所示的35个或35个以上的与年龄相关的表观遗传标记;如表S5所示的40个或40个以上的与年龄相关的表观遗传标记;如表S5所示的45个或45个以上的与年龄相关的表观遗传标记或如表S5所示的50个或50个以上的与年龄相关的表观遗传标记。

18. 权利要求15的方法,其中,实足年龄与预测年龄的关联至少为90%。

19. 权利要求13的方法,其中,在一个受试者基因组中,受胞嘧啶C-5位甲基化约束的具有单个CpG残留物的一套标记包括一个或多个甲基化标记为:表格S4的cg05652533,表格S4的cg27367526,表格S4的cg18404041,图9、表格S3和S5的cg23606718,图9、表格S3和S5的cg16867657,在位置52008487染色体3上的cg04474832,在位置21369010染色体22上的cg05442902,在位置25652602染色体6上的cg06493994,在位置236557682染色体1上的cg09809672,在位置30849114染色体12上的cg19722847,在位置18122719染色体6上的cg22736354,表格S4的cg05652533,表格S4的cg27367526,表格S4的cg18404041,在位置131513927染色体2上的cg23606718,和/或在位置11044877染色体6上的cg16867657。

20. 权利要求13的方法,其中,在一个受试者基因组中,受胞嘧啶C-5位甲基化约束的具有单个CpG残留物的一套标记包括甲基化标记为:表格S4的cg05652533,表格S4的cg27367526,表格S4的cg18404041,图9、表格S3和S5的cg23606718,图9、表格S3和S5的cg16867657,在位置52008487染色体3上的cg04474832,在位置21369010染色体22上的cg05442902,在位置25652602染色体6上的cg06493994,在位置236557682染色体1上的cg09809672,在位置30849114染色体12上的cg19722847,在位置18122719染色体6上的cg22736354,表格S4的cg05652533,表格S4的cg27367526,表格S4的cg18404041,在位置

131513927染色体2上的cg23606718,和/或在位置11044877染色体6上的cg16867657。

21. 权利要求13的方法,其中,在一个受试者基因组中,受胞嘧啶C-5位甲基化约束的具有单个CpG残留物的一套标记包括甲基化标记为:位置52008487上染色体3的cg04474832,位置21369010上染色体22的cg05442902,位置25652602上染色体6的cg06493994,位置236557682上染色体1的cg09809672,位置上30849114染色体12的cg19722847和位置18122719上染色体6的cg22736354。

22. 权利要求13的方法,其中,在一个受试者基因组中,受胞嘧啶C-5位甲基化约束的具有单个CpG残留物的一套标记包括甲基化标记为:图9或表S3的cg20822990,图9或表S3的cg04400972,图9或表S3的cg16054275,图9或表S3的cg03607117,图9或表S3的cg20052760,图9或表S3的cg16867657,图9或表S3的cg06493994,图9或表S3的cg06685111,图9或表S3的cg00486113,图9或表S3的cg20426994,图9或表S3的cg14361627,图9或表S3的cg08097417,图9或表S3的cg07955995,图9或表S3的cg22285878和/或图9或表S3的cg08540945。

23. 一种用于诊断受试者肿瘤存在的方法,包括:

(a) 获得疑似含有肿瘤的生物样品及已知不含有肿瘤的同类型第二份生物样品;

(b) 根据权利要求2的方法预测各个生物样品的年龄;

(c) 对比两种样品所预测的年龄,得到具有肿瘤的生物样品比无肿瘤的生物样品预测年龄较年长。

24. 权利要求13的方法,其中,具有肿瘤的生物样品血液、淋巴结、肝、脑、食道、气管、胃、肠、胰腺、喉、舌、骨、卵巢、子宫、子宫颈、腹膜、前列腺、睾丸、乳腺、肾、肺或皮肤中得到。

25. 权利要求13的方法,其中,在受试者表观基因组中与年龄相关的一组表观遗传标记包括图9、表S3和S5的cg23606718和/或图9、表S3和S5的cg16867657的甲基化标记的任何一种。

26. 权利要求13的方法,其中,具有肿瘤细胞的生物样品可预测的较年长的预测年龄至少比无肿瘤细胞的生物样品多30%。

27. 权利要求13的方法,其中,在受试者表观基因组中与年龄相关的一套表观遗传标记包括图9、表S3和S5的cg23606718和/或图9、表S3和S5的cg16867657的甲基化标记的任何一种及具有肿瘤细胞的生物样品可预测的较年长的预测年龄至少比无肿瘤细胞的生物样品多30%。

28. 权利要求22的方法,其中,在受试者基因组的大多数与年龄相关的表观遗传标记在具有肿瘤细胞的生物样品比在无肿瘤细胞同样类型的生物样品中可预测较年长的年龄。

29. 权利要求28的方法,其中,在受试者基因组中的大多数与年龄相关的表观遗传标记在具有肿瘤细胞的生物样品比在无肿瘤细胞同样类型的生物样品中预测的较年长的年龄可超过总的与年龄相关的表观遗传标记的70%。

30. 权利要求2的方法,其中,一种或多种可与未修饰的胞嘧啶和5-甲基-修饰的胞嘧啶进行不同的反应的探针选自于包括亚硫酸氢钠、焦亚硫酸钠和/或亚硫酸氢盐的一组。

31. 权利要求2的方法,其中,分离的基因组DNA或核DNA与一种或多种探针反应的结果是未修饰胞嘧啶脱氨成尿嘧啶和未改变的5-甲基胞嘧啶。

32. 权利要求2的方法,其中,分析甲基化状态包含(a)DNA扩增和核酸序列的测定及检测与年龄相关的表观遗传标记中CpG二核苷酸C位或是胞嘧啶或是胸腺嘧啶的存在;或者

(b)DNA扩增之后是噬菌体RNA聚合酶的转录、核糖核酸酶裂解和核糖核酸酶裂解产物的基质辅助激光解吸电离飞行时间质谱(MALDI-TOF MS)。

33. 权利要求32的方法,其中,核酸序列的测定涉及到一个或多个下列程序:核酸片段化、限制性内切酶消化、核酸杂交、引物延伸、焦磷酸测序、单核苷酸延伸、具有生物素标记的ddNTP的单核苷酸延伸、具有2,4-二硝基苯酚(DNP)标记的ddNTP的单核苷酸延伸、放射性同位素标记、非放射性标记掺入、荧光标记掺入、生物素掺入、抗原-抗体复合物的形成、抗体检测、比色检测、荧光检测、用荧光染料标记的抗体的检测、标记的抗生物素蛋白或抗生物素链菌素的检测、磁珠分析或检测方法、信号放大、聚合酶链反应、具有热稳定DNA聚合酶的DNA扩增、phi-29DNA聚合酶DNA扩增、RNA的合成、体外转录、噬菌体RNA聚合酶转录、T7RNA聚合酶的转录、SP6RNA聚合酶的转录、T3RNA聚合酶转录、RNA酶消化、RNA酶A消化、DNA克隆、细菌转化、凝胶电泳、质谱法、MALDI-TOF质谱法、微阵列分析、荧光扫描分析、自动化的数字图像捕捉、自动化的数字图像分析、比率分析和Infinium® HumanMethylation450磁珠芯片分析。

34. 权利要求2的方法,其中,最初呈现在每个与年龄相关的表观遗传标记的5-甲基胞嘧啶和胞嘧啶的比可以是暴露于一种或多种探针后CpG二核苷酸嘧啶位置上的胞嘧啶与胸腺嘧啶的比,随后进行核酸扩增产品的分析。

35. 权利要求2的方法,其中,最初呈现在每个与年龄相关的表观遗传标记中的未修饰的胞嘧啶和5-甲基胞嘧啶的比可以是暴露于一种或多种探针后CpG二核苷酸嘧啶位置上的胸腺嘧啶与胞嘧啶的比,随后进行核酸扩增产品的分析。

36. 权利要求3的方法,其中,一种或多种可鉴别特定DNA序列的限制性内切酶探针包含至少一个CpG二核苷酸及基于胞嘧啶的C-5位位置甲基基团的存在或不存在,DNA裂解的方式不同,所述的CpG二核苷酸选自于限制性内切酶,所述的限制性内切酶选自于包括AatII、Acc65I、AccI、AciI、AcII、AfeI、AgeI、AhdI、AieI、ApaI、ApaLI、AscI、AsiSI、AvaI、AvaII、BaeI、BanI、BbrPI、BbvCI、BceAI、BcgI、BcoDI、BfuAI、BfuCI、BglI、BmgBI、BsaAI、BsaBI、BsaHI、BsaI、BseYI、BsiEI、BsiWI、BslI、BsmAI、BsmBI、BsmFI、BspDI、BsrBI、BsrFI、BssHI、BssKI、BstAPI、BstBI、BstUI、BstZ17I、Cac8I、ClaI、DpnI、DraIII、DrdI、EaeI、EagI、EarI、EciI、Eco53kI、EcoRI、EcoRV、FauI、Fnu4HI、FokI、FseI、FspI、HaeII、HaeIII、HgaI、HhaI、HincII、HinfI、HinPII、HpaI、HpaII、Hpy166II、Hpy188III、Hpy99I、HpyAV、HpyCH4IV、HpyCH4V、KasI、MboI、MluI、MmeI、MspAII、MwoI、NaeI、NarI、NciI、NgoMIV、NheI、NlaIV、NotI、NruI、Nt.BbvCI、Nt.BsmAI、Nt.CviPII、PaeR7I、PhoI、PleI、PfuTI、PmeI、PmlI、PshAI、PspOMI、PspXI、PvuI、RsaI、RsrII、SacII、SalI、Sau3AI、Sau96I、ScrFI、SfaNI、SfiI、SfoI、SgrAI、SmaI、SnaBI、StyD4I、TfiI、TliI、TseI、TspMI、XhoI、XmaI,和ZraI的一组。

37. 权利要求3的方法,其中,分离的基因组DNA或核DNA与一种或多种限制性内切酶探针反应的结果是当CpG二核苷酸中的胞嘧啶未修饰时在限制性内切酶分裂位点产生一个双链DNA断裂或者是当CpG二核苷酸中的胞嘧啶在其C-5位用甲基基团修饰时在限制性内切酶分裂位点无DNA双链断裂。

38. 权利要求3的方法,其中,在步骤(c)中分析甲基化状态包括DNA扩增和(a)由于限制性内切酶的分裂而特定DNA端或片段及(b)与特定年龄相关的表观遗传标记相关联的完整限制性内切酶分裂位点的与年龄相关的表观遗传标记的分析。

39. 权利要求3的方法,其中,最初呈现在每个与年龄相关的表观遗传标记未修饰的5-甲基胞嘧啶的比例可以为由于限制性内切酶的耐分裂具有完整限制性内切酶分裂位点与年龄相关的表观遗传标记的分数或者百分比。

40. 权利要求3的方法,其中,最初呈现在每个与年龄相关的表观遗传标记未修饰的胞嘧啶的比例可以为限制性内切酶的分裂的与年龄相关的表观遗传标记的分数或者百分比。

41. 权利要求3的方法,其中,最初呈现在每个与年龄相关的表观遗传标记的5-甲基胞嘧啶与未修饰的胞嘧啶的比可以为完整地限制性内切酶分裂位点的数目或浓度与由与年龄相关的表观遗传标记的限制性内切酶所产生的DNA双链断裂数目或浓度之比。

42. 权利要求3的方法,其中,最初呈现在每个与年龄相关的表观遗传标记的未修饰的胞嘧啶与5-甲基胞嘧啶的比可以为由限制性内切酶所产生的DNA双链断裂的数目或浓度与由于与年龄相关的表观遗传标记的5-甲基胞嘧啶的存在抗限制性内切酶分裂的限制性内切酶分裂位点的数目或浓度之比。

43. 权利要求1的方法,其中在(b)中,确定甲基化状态包括下列步骤:

(a) 从样品中分离基因组DNA或核DNA;

(b) 用一种或多种探针探测分离的基因组DNA或核DNA,所述的一种或多种探针可与未修饰的胞嘧啶和5-甲基-修饰的胞嘧啶进行不同的反应;

(c) 扩增步骤(b)中的DNA;

(d) 用一种或多种限制性内切酶消化步骤(c)扩增的DNA,所述的一种或多种限制性内切酶鉴别含有CpG二核苷酸的限制性内切酶位点,但不能消化从CpG二核苷酸突变为TpG二核苷酸的限制性内切酶位点;

(e) 基于敏感或耐消化的在步骤(d)中限制性内切酶位点的分数或百分比确定最初呈现在每个与年龄相关的表观遗传标记的5-甲基胞嘧啶或胞嘧啶的比例;或者基于在步骤(d)中敏感性消化限制性内切酶位点的数目或浓度与耐消化限制性内切酶位点的数目或浓度之比,确定最初呈现在每个与年龄相关的表观遗传标记的5-甲基胞嘧啶与未修饰的胞嘧啶的比;基于在步骤(d)中敏感性消化限制性内切酶位点的数目或浓度与耐消化限制性内切酶位点的数目或浓度之比,确定最初呈现在每个与年龄相关的表观遗传标记的未修饰的胞嘧啶与5-甲基胞嘧啶的比。

44. 权利要求1的方法,其中在(b)中,确定选自图9、表S3、S4和/或S5中受试者表观基因组的与年龄相关的一组表观遗传标记的甲基化状态包括下列步骤:

(a) 分离基因组DNA或核DNA;

(b) 破碎基因组DNA或核DNA;

(c) 把步骤(b)中破碎的DNA暴露于5-甲基胞嘧啶结合蛋白中;

(d) 从步骤(c)中5-甲基胞嘧啶结合蛋白-自由DNA片段中分离5-甲基胞嘧啶结合蛋白结合的DNA片段;

(e) 通过确定每一个与年龄相关的表观遗传标记含有DNA片段的5-甲基胞嘧啶结合蛋白结合的或游离的DNA片段的分数或百分比,确定每个与年龄相关的表观遗传标记的含DNA片段的5-甲基胞嘧啶或含DNA片段的未修饰的胞嘧啶的比例;或者通过确定5-甲基胞嘧啶结合蛋白结合的DNA片段的数目或浓度与5-甲基胞嘧啶结合蛋白-自由DNA片段的数目或浓度的比,确定每个与年龄相关的表观遗传标记的含DNA片段的5-甲基胞嘧啶与含DNA片段的

未修饰的胞嘧啶的比例;或者通过确定5-甲基胞嘧啶结合蛋白-自由DNA片段的数目或浓度与5-甲基胞嘧啶结合蛋白结合的DNA片段的数目或浓度的比,确定每个与年龄相关的表观遗传标记的含DNA片段的未修饰的胞嘧啶与含DNA片段的5-甲基胞嘧啶的比例。

45. 权利要求44的方法,其中,5-甲基胞嘧啶结合蛋白可以是5-甲基胞嘧啶、MeCP2、MBD2、MBD2/MBD3L1复合物、MBD2的核MBD区域、或者聚MBD蛋白、天然存在的5-甲基胞嘧啶结合蛋白、遗传工程5-甲基胞嘧啶结合蛋白、或其衍生物或其片段的抗体。

46. 权利要求45的方法,其中,从5-甲基胞嘧啶结合蛋白-自由DNA片段中分离5-甲基胞嘧啶结合蛋白-结合DNA片段包括免疫沉淀法、免疫捕获法、固相色谱法、液相色谱法和/或凝胶电泳法。

47. 一种提用于确定受试者表观甲基化衰老速率(AMAR)的方法,包括:

(a)用权利要求1的方法预测年龄;及

(b)用(a)中预测的年龄除以实际年龄,从而确定受试者表观甲基化衰老速率。

48. 权利要求1的方法,其中,对比与年龄相关的受试者的一套表观遗传标记的甲基化状态与和年龄相关的参考人群的相同标记的甲基化状态,包括一种统计方法选自于由多元回归方法、线性回归分析、表格方法或者包括弹性网络的图形化方法、套索回归法、岭回归法、最小二乘拟合法、二项检验法、Shapiro-Wilk检验法、Grubb统计法、Benjamini-Hochberg FDR、方差分析、熵统计法和/或Shannon熵构成的组合。

49. 一种从人类组织通过基于权利要求1方法的表观遗传标记预测受试者的年龄的人类实际年龄的法医诊断。

50. 通过基于权利要求方法1的受试者表观遗传组预测受试者的年龄的受试者健康评估的方法。

51. 一种筛选有益试剂是延缓或加速老化过程的方法,包括:

(a)从活有机体中获得生物样品,及所述的有机体优选为从活有机体衍生的培养细胞、组织或器官;

(b)预测年龄或有机体的AMAR,随权利要求1或46的方法之后,使用适当的与年龄相关的表观遗传标记的有机体,以便由权利要求1所提供的人类受试者与年龄相关的表观遗传标记可以需要被已检查过的有机体的与年龄相关的表观遗传标记所取代;

(c)把有机体或步骤(a)中来自活有机体的培养活细胞、组织或器官暴露于单剂量、多剂量或连续剂量的有益试剂中;从而从活有机体或培养活细胞、组织或器官中获得生物样品

(d)从活有机体或步骤(c)中培养活细胞、组织或器官中获得生物样品;

(e)根据权利要求1或47的方法,使用适当的与年龄相关的表观遗传标记的有机体预测年龄或步骤(d)中生物样品的有机体的AMAR,以便由权利要求1所提供的人类受试者的与年龄相关的表观遗传标记可以需要被已检查过的有机体的与年龄相关的表观遗传标记所取代;

(f)在另一个个体上实施步骤(a)-(e)的相同的步骤,该个体来自于同样的生物或完全相同的培养的活细胞、组织或者同一个体的器官或者未用有益试剂处理或在步骤(c)中用安慰剂处理过的有机体,及

(g)比较步骤(b)中相同预测年龄或AMAR的生物样品,步骤(e)中预测年龄或AMAR与步

骤(f)中预测年龄或AMAR,得到步骤(e)中的一个较低的值或值的范围表明有益试剂可减缓衰老过程而较高的值或值的范围表明有益试剂可加速衰老过程;从而筛选有益试剂是延缓或加速老化过程。

52. 权利要求51的方法,其中,有益试剂包括抗氧化剂、还原剂、DNA损伤剂、维生素、营养补充剂、食品、食品添加剂、食品着色剂、盐、蔬菜、蔬菜提取物、水果、水果提取物、花、花提取物、香料、种子、籽提取物、香草、草本提取物、植物提取物、纤维、脂肪、脂肪酸、油、糖、人造甜味剂、益生菌、乙醇、葡萄酒、真菌、霉菌、奶油、洗剂、粉末、化妆品、防晒剂、气体、污染物、烟雾、环境污染物、油漆、溶剂、有机溶剂、塑料、增塑剂、双酚、酚类化合物、烟草、吸入剂、药物、生物制品、激素、内分泌干扰物、环境雌激素、激素拮抗剂、激素激动剂,咖啡因、植物雌激素、金属、酶、螯合剂、酸奶、硫化合物,物理屏障、电磁障碍和辐射屏障。

53. 权利要求51的方法,其中,有机体是酵母、果蝇、鱼类、蠕虫、昆虫、斑马鱼、线虫、植物或哺乳动物。

54. 权利要求53的方法,其中,哺乳动物为人类、鼠、猴、猫、犬、马、牛、猪、绵羊、山羊、兔、哺乳类家畜、哺乳类运动动物和哺乳类宠物。

55. 一种用于鉴定已知年龄受试者的生物样品的组织类型的方法,包含:

- (a) 确定受试者的实足年龄;
- (b) 通过权利要求1生物方法确定来自于生物样品的受试者的预测年龄;
- (c) 对比各种组织类型的相关预测年龄的参考标准与实足年龄;
- (d) 确定步骤(c)中各种组织类型参考标准中与预测年龄来自于(b)中最匹配的值;及
- (e) 基于步骤(d)中最匹配的值确定生物样品的组织类型,鉴别已知年龄受试者的生物样品的组织类型。

56. 一种用于鉴定已知年龄受试者的生物样品的组织类型的方法,包括:

- (a) 确定受试者的实际年龄;
- (b) 通过从受试者实足年龄分出受试者预测年龄,确定来自于生物样品的受试者的AMAR;
- (c) 对比相关AMAR的参考标准与各种组织类型的实足年龄;
- (d) 确定步骤(c)中各种组织类型参考标准的AMAR来自于步骤(b)最匹配的值;及
- (e) 基于步骤(d)中最匹配的值确定生物样品的组织类型,从而定已知年龄受试者的生物样品的组织类型。

57. 权利要求55或56的方法,其中,受试者基因组中与年龄相关的一套表观遗传标记包含任何一种或多种图9、表S3和S5的cg23606718和/或图9、表S3和S5的cg16867657表观遗传标记。

58. 一种基于表观遗传修饰影响基因表达预测受试者年龄的方法,包括:

- (a) 获得受试者的生物样品;
- (b) 确定与年龄相关的表观遗传标记相关联的其表达随着年龄变化的一个或多个的基因的表达;
- (c) 对比与年龄相关的表观遗传标记相关联的其表达随着年龄变化的一个或多个的基因的表达与来自与年龄相关的参考人群的相同基因的表达;及
- (d) 获得受试者预测年龄的一个值或值的范围;其中,对比与年龄相关的表观遗传标记

相关联的其表达随着年龄变化的一个或多个的基因的表达与来自与年龄相关的参考人群的同基因的表达可以包括使用任一统计方法、多变量回归分析法、线性回归分析、表格法、或图解法基于与年龄相关的表观遗传标记相关联的其表达随着年龄变化的基因的表达预测受试者的年龄；从而基于表观遗传修饰影响基因表达预测受试者的年龄。

59. 权利要求58的方法，其中，与年龄相关的表观遗传标记相关联的其表达随着年龄变化的一个或多个的基因的表达包括一个或多个列在表S6或表S7的基因。

60. 权利要求58的方法，其中，基因表达是转录或翻译。

61. 权利要求60的方法，其中，转录导致RNA转录体的产生，翻译导致蛋白质的产生。

62. 一种用于筛选受试者组织样品以基于由权利要求1方法中的受试者的表观基因组预测组织样品年龄的方法。

63. 权利要求62的方法，其中，在高通量的筛选试验中组织样品暴露于至少一种测试试剂。

64. 权利要求62的方法，其中，所述过程用于任一诊断和/或高通量筛选。

65. 权利要求48的方法，其中，统计方法是多元回归法或线性回归法。

66. 一种基于受试者的组织或器官的表观基因组预测受试者的组织或器官年龄的方法，包括：

(a) 从受试者中获得组织或者器官的生物样品；

(b) 在选自于图9、表S3、S4和/或S5受试者表观基因组中确定与年龄相关的一套表观遗传标记的甲基化状态；及

(c) 对比受试者与年龄相关的一套表观遗传标记的甲基化状态与来自与年龄相关的参考人群相同标记的甲基化状态，以获得预测组织或器官年龄的一个值或值的范围，从而基于受试者的组织或者器官的表观基因组预测受试者组织或者器官的年龄。

67. 权利要求66的方法，其中，来自与年龄相关的参考人群的同标记的甲基化状态在相同类型的组织或器官中确定。

68. 权利要求66的方法，其中，来自与年龄相关的参考人群的同标记的甲基化状态在不同类型的组织或器官中确定。

69. 权利要求66的方法，其中，来自与年龄相关的参考人群同标记的甲基化状态在血液或分离血液中确定。

70. 一种用于确定受试者组织或器官不同衰老速率的方法，包括：

(a) 从受试者不同组织或器官中获得生物样品；

(b) 使用权利要求66的方法预测组织或器官的年龄；

(c) 对比步骤(b)所确定的预测年龄，其中预测年龄的差别表明受试者组织或器官衰老速率的差别，从而确定受试者组织或器官不同老化速率。

71. 权利要求70的方法，其中，预测年龄可除以受试者的实足年龄以获得AMAR。

72. 权利要求1的方法，其中，所述方法为自动化方法。

73. 权利要求1的方法，其中，受试者是人、哺乳动物、动物、植物或任何多细胞有机体。

74. 一种基于受试者遗传物质的表观遗传修饰确定受试者年龄的试剂盒，包括与年龄相关的一套表观遗传标记或图9、表S3、表S4或表S5所列举的标记。

75. 一种基于受试者的表观遗传标记确定受试者年龄的试剂盒，使用图9、表S3、S4和/

或S5所提供的与年龄相关的一套表观遗传标记。

76. 权利要求75中的试剂盒,其中,与年龄相关的表观遗传标记可包括具有CpG二核苷酸的核酸。

77. 权利要求75中的试剂盒,其中,所述CpG二核苷酸的胞嘧啶在C-5位的甲基化受年龄依赖性变化的影响。

78. 权利要求75中的试剂盒,其中,所述CpG二核苷酸在如图9、表S3、S4和/或S5所描述的染色体位置。

79. 权利要求75中的试剂盒,其中,与年龄相关的表观遗传标记为人类的标记,并选自于位置52008487上染色体3的cg04474832、位置21369010上染色体22的cg05442902、位置25652602上染色体6的cg06493994、位置236557682上染色体1的cg09809672、位置30849114上染色体12的cg19722847、位置18122719上染色体6的cg22736354、表S4的cg05652533、表S4的cg27367526、表S4的cg18404041、位置131513927上染色体2的cg23606718和位置11044877上染色体6的cg16867657。

80. 权利要求75中的试剂盒,其中,与年龄相关的表观遗传标记具有图9所提供的序列。

用于预测年龄和鉴别诱发或者抑制早衰试剂的方法

[0001] 本发明是在政府支持下,根据NIH授予的P50GM085764和R01E5014811的授权号及NEI/NIH授予的EY014428、EY018660、EY019270和EY021374的授权号进行的。政府对本发明拥有一定权利。

[0002] 贯穿本申请中引用了各种公开文件。这些公开文件全部的公开内容通过引用并入本申请中,以便更充分地描述本发明涉及的本领域的状态。

[0003] 发明背景

[0004] 并不是每个人以同样的方式衰老。众所周知,女人往往比男人更长寿,及生活方式的选择如吸烟和体能可加速或者延缓衰老过程(Steven N.,2006;Blair et al.,1989)。这些观察资料已导致寻找可用于预测、监测以及提供使人深入了解年龄相关的生理性下降和疾病的年龄分子标记。一种此类的标记是端粒长度,与年龄密切相关的分子特征(Harley et al.,1990),其已经表明衰变在环境压力下具有加速的速率(EpeI et al.,2004;Valdes et al.)。另一个标记为基因表达,尤其是在新陈代谢和DNA修复途径中起作用的基因,其通过可跨不同的组织类型和有机体的范围预测年龄(Fraser et al.,2005;Zahn et al.,2007;de Magalhães et al.,2009)。

[0005] 愈来愈多的研究已报道了年龄与表观基因组状态之间的关联-除了主要核苷酸序列变化的DNA修改的设置(Fraga and Esteller,2007)。尤其是,DNA甲基化与长时间范围内的实足年龄相关联(Aiisch et al.,2012;Christensen et al.,2009;Boilati et al.,2009;Boks et al.,2009;Rakyan et al.,2010;Bocklandt et al.,2011;Bell et al.,2012)及甲基化的改变已与复杂的年龄相关性疾病如代谢性疾病(Barres and Zierath,2011)和癌症(Jones and Laird,1999;Esteller,2008)相关联。研究也发现一个被称为“表观遗传学漂移”的现象,由此在同卵双胞胎中的DNA甲基化标记作为年龄的函数差别越来越大(Fraga et al.,2005;Boks et al.,2009)。因此,作为固定标记的表观基因组的想法让位于作为动态展望反映各种实龄变化的表观基因组模型。目前的挑战是确定这些变化是否可被系统地描述和模拟以监测人体衰老的不同速率,及其与相关的临床或者环境变量联系在一起。

[0006] 尽管其归因为至少两个基本因素,但是驱使衰老甲基化的变化机理尚未很好被理解(Vijg and Campisi,2008;Fraga et al.,2005)。首先,环境暴露会随着时间的推移激活与表观基因组一致及可预测变化相关的细胞程序。例如,应力已表明可通过在DNA甲基化中的特定变化以改变基因表达模式(Murgatroyd et al.,2009)。此外,自发的表观遗传变化可在环境应力或无环境应力下发生,从而导致衰老个体之间表观基因组根本不可预测的差异。自发变化可以通过化学试剂破坏DNA甲基引起或者通过DNA复制过程中复制甲基化状态的错误引起。这两种机理导致衰老个体之间的甲基化组的不同,这表明甲基化状态的定量测量可以鉴别涉及衰老速率减缓或加速的因素。

[0007] 为了更好地了解甲基组如何变老及确定人类衰老速率是否可以量化和比较,我们发起了一个项目,以执行跨越广泛年龄范围的大型个体行列的全基因组甲基化分析。基于这些调查结果,我们构建了一个可预测衰老速率的模式,我们表明的衰老速率受性别和特

定基因变异的影响。这些数据帮助解释表观遗传漂移,及表明在甲基化中与年龄相关的变化随着时间的推移导致转录模式的变化。这些调查结果在第二大型行列中被重复。

[0008] 从分子谱中测量人体衰老的能力在许多领域具有实际意义,包括疾病预防和治疗、取证和延长生命。尽管实足年龄与DNA甲基化中的变化有关联,但甲基化组还没用于测量和比较人体衰老速率。在此,我们创建了一种衰老的定量模型,用于从656个年龄从19岁到101岁的人体个体的全血中测量超过45万个CpG标记。这种模型测量个体甲基化衰老的速率。进一步地,我们发现,衰老速率的差异可解释表观遗传漂移及在转录中得到反映。我们的发现强调了衰老过程的特定组成,及提供了取证方法、试剂减缓或加速衰老的筛选方法及预防和治疗疾病的方法。

发明概要

[0009] 本发明提供了基于受试者表观基因组预测受试者年龄的方法。一个实施例中,该方法包括(a)获得受试者的生物样品;(b)确定如图9、表S3、S4和/或S5任一所示的在受试者表观基因组中与年龄相关的一套表观遗传标记的甲基化状态;及(c)对比受试者与年龄相关的一套表观遗传标记的甲基化状态和年龄相关的参考人群的相同标记的甲基化状态,以获得用于预测受试者年龄的一个值或者一个值的范围,从而基于受试者的表观基因组预测受试者的年龄。

[0010] 本发明还提供了用于鉴别来自已知实足年龄受试者的生物样品的组织类型的方法。一个实施例中,该方法包括(a)确定受试者的实足年龄;(b)通过将受试者的预测年龄与受试者的实足年龄分开,从生物样品中确定受试者的AMAR;(c)对比与AMAR相关的参考标准与各种组织类型的实足年龄;(d)确定与步骤c中各种组织类型的AMAR参考标准最匹配的步骤b中的数值;及(e)基于步骤d中的最匹配,设定生物样品的组织类型,从而鉴别来自已知实足年龄的受试者的生物样品的组织类型。

[0011] 本发明还提供了基于与年龄相关的表观遗传修饰影响基因表达预测受试者年龄的方法,包括(a)获得受试者的生物样品;(b)确定与年龄相关的其表达随年龄而变化的表观遗传标记的一个或者多个基因的表达;(c)对比一个或者多个与随着年龄的变化与年龄相关的表观遗传标记表达变化相关联的基因表达与来自年龄相关参考人群的相同基因表达;及(d)获得用于预测受试者年龄的一个值或一个值的范围;其中,对比一个或者多个与随着年龄的变化与年龄相关的表观遗传标记表达变化相关联的基因表达与来自年龄相关参考人群的相同基因表达,包括统计方法、多元回归方法、线性回归分析、列表法或者图解法中的任一基于与年龄相关的表观遗传标记相关联的基因表达预测受试者年龄的方法,其表观遗传标记的表达随年龄变化,从而基于与年龄相关的表观遗传修饰影响基因表达预测受试者的年龄。

[0012] 本发明还提供了基于受试者组织或者器官的表观基因组预测受试者组织或者器官的年龄的方法。一个实施例中,该方法包括(a)从受试者获得组织或者器官的生物样品;(b)确定选自于图9、表S3、S4和/或S5受试者表观基因组中与年龄相关的一套表观遗传标记的甲基化状态;及(c)对比受试者的与年龄相关的此套表观遗传标记的甲基化状态与来自年龄相关参考人群的相同标记的甲基化状态,以获得用于预测组织或者器官年龄的一个值或者一个值的范围,从而基于受试者组织或者器官的表观基因组预测受试者组织或者器官

的年龄。

[0013] 本发明还提供了基于受试者的包括任一与年龄有关的表观遗传标记或如图9、表S3、表S4或表S5所列举的标记的遗传物质表观遗传修饰确定受试者年龄的试剂盒。

[0014] 本发明还提供了基于利用图9、表S3、S4和/或S5提供的与年龄有关的一套表观遗传标记的受试者的表观基因组预测受试者年龄的试剂盒。

[0015] 附图简述

[0016] 图1衰老甲基化的全局数据。(A)标记为cg16867657的甲基化分值的密度图,年轻个体(绿色)与年老个体(蓝色)分开。(B)所有个体年龄分布的直方图。(C)与年龄相关的甲基化标记的热图,以关联大小进行排序(回归系数)。个体从最年轻的到最年老的顺序排列。也参见图S1和表S1和S2,分别为与年龄相关的区域特定例子及用于注释重合表。

[0017] 图2模型预测和临床变量。(A)用于产生衰老预测(蓝盒)的数据(绿盒)和分析(红色椭圆形)的流程图。(B)基于衰老模型对比所有个体的预测年龄与实际年龄。(C)验证行列中个体的样本外预测。(D)基于无临床变量的衰老模型每个个体的显著甲基化衰老速率(AMAR)。衰老速率的分布表明男人比女人衰老快。表S3中提供了衰老模型中所使用的标记表格。也参见图S2和S3、表S3和图9。

[0018] 图3甲基化衰老的遗传效应。(A)我们调查了与年龄相关的甲基化标记相关联的基因变异。对应于14meQTLs,挑选八个基因变异用于验证。其中,七个在验证行列中显著,两个表明与AMAR相关联。(B)甲基化标记cg27367526(STEAP2)和年龄之间的趋势图。变异体rs42663(GTPBP10)的状态引起这种关系的偏移。(C)第二个例子cg18404041和rs2230534(ITIH1,NEK4)。也可参见证实遗传相关性表格的表S4。

[0019] 图4多组织支持。(A)通过完整的衰老模型在TCGA上作出的年龄预测控制样品。在实足年龄与预测年龄之间有高的相关性,但是每个组织具有不同的线性截距和斜率。(B)在调整每个组织的截距和斜率后,模型误差与原始全血数据中模型误差相似。图S2中呈现了癌症样品上作出的年龄预测。(C)来自于TCGA的匹配的正常和肿瘤样本上作出的年龄预测。预测作母体组织(胸腺、肾、肺或皮肤)的线性偏移的调整。(D)肿瘤样本在AMAR中表明显著的增加。也参见图S4和表S5。

[0020] 图5甲基化年龄相关性。分数和偏差(A)甲基化分值显示为标记cg24724428。在群体中的任一子集,我们考虑两组甲基化统计:均值和方差。标记方差是平均甲基化变异的一种方式,被定义为个体甲基化分值和预期甲基化分值的平方差。(B)示出平均甲基化随着年龄的变化为标记cg24724428的密度图。年少组合和年老组以顶部和底部的10%为基础。(C)所有标记和年龄之间的甲基化分值的显著相关性的直方图。P值是有符号的,正值代表随着年龄的增加甲基化增加。超过FDR<0.05阈值的标记被分组到极值箱中。(D)示出甲基化变异随着年龄的变化为标记cg24724428的密度图。(E)以与(D)相同的方式,所有标记和年龄之间的甲基化变异的显著相关的直方图。衰老趋势映射到图S3的CpG岛上。也参见图S5。

[0021] 图6. 甲基化谱宽随年龄的趋势。(A)随年龄增加(红色)及随年龄减少(蓝色)的所有甲基化标记的总回归线。最黑的颜色代表中位回归线,边界代表25%和75%的分位点。渐增和渐减标记均趋向中等的甲基化分值。(B)熵衰老速率以与年龄相关的表观遗传标记除以实足年龄的平均Shannon熵计算。这与AMAR有强相关性。

[0022] 图7. 转录衰老模型。(A)我们用在甲基化中示出衰老趋势的基因的mRNA表达数据

构建了一个衰老模型。由于数据集内最近5年间隔年龄的四舍五入,其标准误差(RMSE=7.22年)增加。(B)与甲基化类似,与女性($p < 10^{-4}$)相比,转录表明男性衰老速率是增加的。也可参见表S6和表S7。

[0023] 图8.示出了生物年龄模型。

[0024] 图9是图9A(9A至9A-7)、图9B(9B至9B-7)、图9C(9C至9C-7)、图9D(9D至9D-7)和图9E(9E至9E-7)的集合,其为电子表格,示出了与年龄相关的表观遗传标记,由“cg”前缀后跟一个数字(cg#)命名,相关于表S3,其中,每一个CpG二核苷酸均如此经检查嵌入到每个系列的第三个子面板标题为“Forward_Sequence”的所在列示出的序列内(如,图9x-2,其中“x”为A-E)。例如,把感兴趣的二核苷酸用括号括起来。其他信息可在具有GEO登录号码GPL13534的基因表达综合(GEO)数据库中和彼彼科娃等,基因组学,2011,98:288-95(Bibikova et al.Genomics,2011,98:288-95)中找到。电子表格的各个子面板应如下所示集合:

[0025]

9A	9A-1	9A-2	9A-3	9A-4	9A-5	9A-6	9A-7
9B	9B-1	9B-2	9B-3	9B-4	9B-5	9B-6	9B-7
9C	9C-1	9C-2	9C-3	9C-4	9C-5	9C-6	9C-7
9D	9D-1	9D-2	9D-3	9D-4	9D-5	9D-6	9D-7
9E	9E-1	9E-2	9E-3	9E-4	9E-5	9E-6	9E-7

[0026] 图S1.与图1有关的衰老关系图的一个例子

[0027] 基因4和Half LIM Domains 2(FHL2)的年龄相关水平。强的衰老相关性在基因中心的CpG岛上显示出几个标记(红色:□ $\log_{10}(p\text{-值})$),与内部启动因子(黑色:平均甲基化分值)相一致。

[0028] 图S2.与图2有关的衰老模型应用于海恩(Heyn)等人的数据集

[0029] 我们从海恩等人的数据集中获得甲基化图谱及应用衰老预测模型。我们的模型成功地分离年老样品和年轻样品(黑圈)。另外,我们应用衰老模型至采用亚硫酸氢盐测序测量而不用我们数据的磁珠芯片测量技术的海恩等人的数据集中的三个样品。尽管技术有差别,该模型成功地分离年轻样品、中年样品和年老样品(绿色圆点)。

[0030] 图S3.与图2有关的测量批次相关变量的影响

[0031] 种族和糖尿病性病态的模型协变量与批次变量高度相关,以致于不能确定它们在衰老过程中的影响。然而,我们为子群(A)欧洲人、(B)西班牙人、(C)非糖尿病人和(D)糖尿病患者构建了不同的模型。每个模型均用于预测其互补行列的年龄。结果表明尽管有协变量和/或批次的影响,仍具有强的预测能力。

[0032] 图S4.与图4有关的正常细胞和肿瘤衰老模型预测

[0033] 使用在主行列中鉴别的模型标记,在匹配的正常细胞和肿瘤样品中构建衰老模型。发现通过正常组织预测的肿瘤样品的衰老速率(AMAR)比预期的高(红色,威尔科克斯(Wilcoxon)测试, $P < 10^{-21}$),及通过肿瘤模型预测的正常样品的衰老速率比预期的低(黑色,威尔科克斯(Wilcoxon)测试, $P < 10^{-17}$)。两个衰老速率的分离也非常显著(威尔科克斯(Wilcoxon)测试, $P < 10^{-25}$)。

[0034] 图S5.与图5有关的CpG岛中衰老趋势图

[0035] (A)27,176CpG岛(黑色)的甲基化分值的一个总基因图谱。在相同的区域(绿色)示出了与年龄甲基化分值相关的衰老系数。标明岛和岸的色带代表75%的置信区间。(B)示出甲基化变异(红色)和变异的衰老系数(绿色)的CpG岛图。

[0036] 附图的详细说明

[0037] 定义

[0038] 如本申请中所使用的,生物年龄(bioage)、化学年龄、甲基化年龄和分子年龄是等价的或者是同义的。使用受试者或有机体与年龄相关的一套表观遗传标记确定生物年龄。在本发明中,从特定的CpG二核苷酸及尤其如甲基化状态在C-5位的胞嘧啶的修饰状态的分析确定生物年龄。

[0039] 实足年龄是受试者或有机体的实际年龄。对于动物和人类,实足年龄可以基于从受孕的那一刻起计算的年龄或者基于从出生时间或日期计算的年龄。从受试者或从有机体中获得细胞、组织或者器官的有机体的实足年龄加上细胞、组织或者器官被置于培养期间来确定细胞、组织或器官的实足年龄。另外,在细胞或组织培养的情况下,实足年龄可能与培养的总时间或累计时间或者通道数量有关。

[0040] 如本申请种所使用的,生物样品的术语“组织”可用“细胞”代替,或者反之亦然。

[0041] 如表S3、S4和S5中所提供的在“标记”或“甲基化标记”列下的,图9所提供的在图9x的“ID”或“Name”列下的,所述的图9x中的x是A-E,在本文中用“cg#”命名所讨论的甲基化标记是与年龄相关的表观遗传标记。在标题“Forward_Sequence”及在括号内探测的特定CpG二核苷酸如[CG],图9中提供了本发明中所探测的每个表观遗传标记探针内的特定CpG二核苷酸。所有“cg#”命名的其他序列信息,如本文表S4和S5中,可从具有GEO登录号GPL13534的基因表达综合(GEO)数据库中美国国立卫生研究院生物技术信息国家中心(Bethesda,MD)获得。

[0042] 图9、表S3、S4和S5所提供的甲基化标记用于使用人类甲基化450磁珠芯片的Illumina公司的Infinium甲基化分析仪。然而,基于本发明所提供的每个cg#的序列、同源性或者序列的正态关联,这些与年龄相关的表观遗传标记可用于Infinium甲基化分析系统外的其他分析中。

[0043] 发明方法

[0044] 本发明提供了基于受试者的表观基因组预测受试者年龄的方法。受试者可以是人类、哺乳动物、动物、植物或任何多细胞有机体。适合的哺乳动物的实施例包括但不限于人类、猴、猿、狗、猫、牛、马、羊、猪、兔、小鼠和大鼠。受试者的年龄可为实足年龄、分子年龄、化学年龄、甲基化年龄或生理年龄。表观基因组可为其中DNA可受到表观遗传修饰的脱氧核糖核酸(DNA)。表观遗传修饰可为CpG残留的甲基化。在一个实施例中,甲基化是甲基基团共价键合在胞嘧啶碳-5(C-5)位置上。

[0045] 在一个实施例中,该方法包括获得受试者的生物样品。另外,该方法包括确定选自图9、表S3、S4和/或S5的受试者表观基因组中的与年龄相关的一套表观遗传标记的甲基化状态。更进一步地,该方法包括对比年龄相关的受试者的一套表观遗传标记的甲基化状态与和年龄相关的参考人群的相同标记的甲基化状态,以获得预测受试者年龄的一个值或一个值的范围,从而基于受试者的表观基因组预测受试者的年龄。

[0046] 在一个实施例中,本方法包括使用统计方法对比与年龄相关的受试者的一套表观

遗传标记的甲基化状态与和年龄相关的参考人群的相同标记的甲基化状态。适合的统计方法的实施例包括但不限于多元回归方法、线性回归分析、表格方法或者包括弹性网络的图形化方法、套索回归法、岭回归法、最小二乘拟合法、二项检验法、Shapiro-Wilk检验法、Grubb统计法、Benjamini-Hochberg FDR、方差分析、熵统计法和/或Shannon熵。在优选的实施例中,统计方法包括多变量回归算法或线性回归算法。

[0047] 按照本发明的实际,确定甲基化状态可包括从样品中分离基因组DNA或核DNA,使分离的基因组DNA或核DNA与一种或多种的探针/试剂(如化学探针/试剂)反应,所述探针/试剂与未修饰的胞嘧啶的做出不同的反应,以便胞嘧啶转化为尿嘧啶。该步骤也可包括通过检测胞嘧啶或尿嘧啶的存在,确定或分析在样品分离的基因组DNA或核DNA中的CpG二核苷酸胞嘧啶位置上(此处也可以称为C位置)的甲基化状态。胞嘧啶或尿嘧啶的存在表明在原CpG二核苷酸中分别存在5-甲基胞嘧啶或未经修饰的胞嘧啶。可选择地,通过限制性内切酶的抗分裂可表明在原CpG二核苷酸中存在5-甲基胞嘧啶。对限制性内切酶分裂敏感性可表明在原CpG二核苷酸中存在未经修饰的胞嘧啶。更进一步地,该步骤可进一步包括确定5-甲基胞嘧啶或未经修饰的胞嘧啶最初呈现在各自与年龄相关的表观遗传标记的比例,或者,可选择地,基于探测分离的基因组DNA或核DNA的结果特征,确定最初呈现在各自与年龄相关的表观遗传标记5-甲基胞嘧啶与未经修饰的胞嘧啶的比或者未经修饰的胞嘧啶与5-甲基胞嘧啶的比。

[0048] 按照本发明的实践,确定甲基化状态可包括从样品中分离基因组DNA或核DNA,受到CpG二核苷酸的影响,在内部或接近限制性内切酶的鉴别点或者分裂位点,及基于在所述CpG二核苷酸的胞嘧啶的C-5位位置上存在或不存在甲基基团DNA裂解方式的不同,用一种或多种可鉴别特定DNA序列的限制性内切酶培育分离的基因组DNA或核DNA。该步骤也包括在样品中分离的基因组DNA或核DNA中,通过表明在内部或接近限制性内切酶的鉴别点或者分裂位点的原CpG二核苷酸中存在5-甲基胞嘧啶的限制性内切酶在潜在的分裂位点的抗分裂,确定或分析CpG二核苷酸C位的甲基化状态。限制性内切酶的分裂敏感性可以表明存在未修饰的胞嘧啶。进一步地,该步骤进一步包括确定5-甲基胞嘧啶或未经修饰的胞嘧啶最初呈现在各自与年龄相关的表观遗传标记的比例,或者,可选择地,确定最初呈现在各自与年龄相关的表观遗传标记5-甲基胞嘧啶与未经修饰的胞嘧啶的比或者未经修饰的胞嘧啶与5-甲基胞嘧啶的比。

[0049] 按照本发明的实践,基于选自图9、表S3、S4和/或S5受试者表观基因组的5个或5个以上的与年龄相关的表观遗传标记;选自图9、表S3、S4和/或S5受试者表观基因组的10个或10个以上的与年龄相关的表观遗传标记;选自图9、表S3、S4和/或S5受试者表观基因组的15个或15个以上的与年龄相关的表观遗传标记;选自图9、表S3、S4和/或S5受试者表观基因组的20个或20个以上的与年龄相关的表观遗传标记;选自图9、表S3、S4和/或S5受试者表观基因组的25个或25个以上的与年龄相关的表观遗传标记;选自图9、表S3、S4和/或S5受试者表观基因组的30个或30个以上的与年龄相关的表观遗传标记;选自图9、表S3、S4和/或S5受试者表观基因组的35个或35个以上的与年龄相关的表观遗传标记;选自图9、表S3、S4和/或S5受试者表观基因组的40个或40个以上的与年龄相关的表观遗传标记;选自图9、表S3、S4和/或S5受试者表观基因组的45个或45个以上的与年龄相关的表观遗传标记;选自图9、表S3、S4和/或S5受试者表观基因组的50个或50个以上的与年龄相关的表观遗传标记;选自图9、

表S3、S4和/或S5受试者表观基因组的55个或55个以上的与年龄相关的表观遗传标记;选自图9、表S3、S4和/或S5受试者表观基因组的60个或60个以上的与年龄相关的表观遗传标记;选自图9、表S3、S4和/或S5受试者表观基因组的65个或65个以上的与年龄相关的表观遗传标记;选自图9、表S3、S4和/或S5受试者表观基因组的70个或70个以上的与年龄相关的表观遗传标记,可确定甲基化状态。

[0050] 进一步地,在优选的实施例中,基于图9或表S3中受试者表观基因组的5个或5个以上的与年龄相关的表观遗传标记;图9或表S3受试者表观基因组的10个或10个以上的与年龄相关的表观遗传标记;图9或表S3受试者表观基因组的15个或15个以上的与年龄相关的表观遗传标记;图9或表S3受试者表观基因组的20个或20个以上的与年龄相关的表观遗传标记;图9或表S3受试者表观基因组的25个或25个以上的与年龄相关的表观遗传标记;图9或表S3受试者表观基因组的30个或30个以上的与年龄相关的表观遗传标记;图9或表S3受试者表观基因组的35个或35个以上的与年龄相关的表观遗传标记;图9或表S3受试者表观基因组的40个或40个以上的与年龄相关的表观遗传标记;图9或表S3受试者表观基因组的45个或45个以上的与年龄相关的表观遗传标记;图9或表S3受试者表观基因组的50个或50个以上的与年龄相关的表观遗传标记;图9或表S3受试者表观基因组的55个或55个以上的与年龄相关的表观遗传标记;图9或表S3受试者表观基因组的60个或60个以上的与年龄相关的表观遗传标记;图9或表S3受试者表观基因组的65个或65个以上的与年龄相关的表观遗传标记,或者图9或表S3受试者表观基因组的70个或70个以上的与年龄相关的表观遗传标记,可确定甲基化状态。例如,在一个受试者基因组中,受胞嘧啶C-5位甲基化约束的具有单个CpG残留物的该套标记可包括一个或多个以下甲基化标记:表S4的cg05652533,表S4的cg27367526,表S4的cg18404041,图9、表S3和S5的cg23606718,图9、表S3和S5的cg16867657,位置52008487上染色体3的cg04474832,位置21369010上染色体22的cg05442902,位置25652602上染色体6的cg06493994,位置236557682上染色体1的cg09809672,位置30849114上染色体12的cg19722847,位置18122719上染色体6的cg22736354,表S4的cg05652533,表S4的cg27367526,表S4的cg18404041,位置上131513927染色体2的cg23606718,和/或位置11044877上染色体6的cg16867657。

[0051] 在一个实施例中,在一个受试者基因组中,受尿嘧啶C-5位甲基化约束的具有单个CpG残留物的该套标记可包括甲基化标记位置52008487上染色体3的cg04474832,位置21369010上染色体22的cg05442902,位置25652602上染色体6的cg06493994,位置236557682上染色体1的cg09809672,位置上30849114染色体12的cg19722847和位置18122719上染色体6的cg22736354。

[0052] 在另一个实施例中,在一个受试者基因组中,受尿嘧啶C-5位甲基化约束的具有单个CpG残留物的该套标记可为任何一个或多个甲基化标记图9或表S3的cg20822990,图9或表S3的cg04400972,图9或表S3的cg16054275,图9或表S3的cg03607117,图9或表S3的cg20052760,图9或表S3的cg16867657,图9或表S3的cg06493994,图9或表S3的cg06685111,图9或表S3的cg00486113,图9或表S3的cg20426994,图9或表S3的cg14361627,图9或表S3的cg08097417,图9或表S3的cg07955995,图9或表S3的cg22285878和/或图9或表S3的cg08540945。

[0053] 在进一步的实施例中,与年龄相关的一套表观遗传标记可为任何一个或多个甲基

化标记图9、表S3和S5的cg23606718和/或图9、表S3和S5的cg16867657。

[0054] 按照本发明的实际,本发明的方法可为自动化方法。

[0055] 按照本发明的实际,生物样本可为血液、淋巴细胞、单核细胞、中性粒细胞、嗜碱性粒细胞、嗜酸性粒细胞、髓系细胞、淋巴系细胞、骨髓、唾液、腮拭试、鼻拭子、尿、粪便、毛、乳腺组织、卵巢组织、子宫组织、宫颈组织、前列腺组织、睾丸组织、脑组织、神经细胞、星形胶质细胞、肝组织、肾、甲状腺组织、胃组织、肠组织、胰腺组织、血管组织、皮肤、肺组织、骨组织、软骨、韧带、肌腱、脂肪细胞、肌细胞、神经元、星形胶质细胞、具有不同传代次数的培养细胞、癌症/肿瘤细胞、癌症/肿瘤组织、正常细胞、正常组织、含有遗传物质核的任一组织或细胞或受试者的已知或未知DNA的形式的遗传物质中的任何一种。

[0056] 肿瘤或癌症细胞可从血液、淋巴结、肝、脑、食道、气管、胃、肠、胰腺、喉、舌、骨、卵巢、子宫、子宫颈、腹膜、前列腺、睾丸、乳腺、肾、肺或皮肤中得到。具有肿瘤或癌症细胞的生物样品可预测有较年长的预测年龄,至少比无肿瘤或癌症细胞的生物样品长约30%或40%。

[0057] 在一个实施例中,与年龄相关的表观遗传标记包括一个CpG残留物。胞嘧啶C-5位的甲基化可能与受试者相关的物种的实足年龄不同。例如,与受试者相关的物种可以是人类。

[0058] 在另一个实施例中,与年龄相关的一套表观遗传标记可包括在受试者基因组中受胞嘧啶C-5位年龄依赖性甲基化的影响的个体CpG残留物。该组标记可包括约70个含有与年龄相关的表观遗传标记独特的CpG残留物。此外,该套标记可包括一个或多个如图9、表S3、表S4或表S5所示的标记。

[0059] 例如,与年龄相关的一套标记可包括如图9或表S3所示的5个或5个以上的与年龄相关的表观遗传标记;如图9或表S3所示的10个或10个以上的与年龄相关的表观遗传标记;如图9或表S3所示的15个或15个以上的与年龄相关的表观遗传标记;如图9或表S3所示的20个或20个以上的与年龄相关的表观遗传标记;如图9或表S3所示的25个或25个以上的与年龄相关的表观遗传标记;如图9或表S3所示的30个或30个以上的与年龄相关的表观遗传标记;如图9或表S3所示的35个或35个以上的与年龄相关的表观遗传标记;如图9或表S3所示的40个或40个以上的与年龄相关的表观遗传标记;如图9或表S3所示的45个或45个以上的与年龄相关的表观遗传标记;如图9或表S3所示的50个或50个以上的与年龄相关的表观遗传标记;如图9或表S3所示的55个或55个以上的与年龄相关的表观遗传标记;如图9或表S3所示的60个或60个以上的与年龄相关的表观遗传标记;如图9或表S3所示的65个或65个以上的与年龄相关的表观遗传标记或者如图9或表S3所示的70个或70个以上的与年龄相关的表观遗传标记。

[0060] 在另一个实施例中,与年龄相关的一组标记可包括如表格S5所示的5个或5个以上的与年龄相关的表观遗传标记;如表S5所示的10个或10个以上的与年龄相关的表观遗传标记;如表S5所示的15个或15个以上的与年龄相关的表观遗传标记;如表S5所示的20个或20个以上的与年龄相关的表观遗传标记;如表S5所示的25个或25个以上的与年龄相关的表观遗传标记;如表S5所示的30个或30个以上的与年龄相关的表观遗传标记;如表S5所示的35个或35个以上的与年龄相关的表观遗传标记;如表S5所示的40个或40个以上的与年龄相关的表观遗传标记;如表S5所示的45个或45个以上的与年龄相关的表观遗传标记或如表S5所

示的50个或50个以上的与年龄相关的表观遗传标记。

[0061] 仅仅通过举例的方式,实足年龄与预测年龄的关联可至少约为80%、90%或91%,具有低于约5年的误差。

[0062] 在另一个实施例中,与年龄相关的一套表观遗传标记可为图9、表S3和S5的cg23606718和/或图9、表S3和S5的cg16867657中的任一甲基化标记。具有肿瘤细胞或癌症细胞的生物样品可预测的较年长的预测年龄,至少比无肿瘤细胞或癌症细胞的生物样品长约30%或40%。

[0063] 在一个实施例中,在受试者基因组的大多数与年龄相关的表观遗传标记在具有肿瘤细胞的生物样品比在无肿瘤细胞同样类型的生物样品中可预测较年长的年龄。类似地,癌前病变比无此类病变的正常组织类型可显示较年长的生物年龄或预测年龄。

[0064] 在另一个实施例中,在受试者基因组的大多数与年龄相关的表观遗传标记在具有肿瘤细胞的生物样品比在无肿瘤细胞同样类型的生物样品中预测的较年长的年龄可超过总的与年龄相关的表观遗传标记的约70%。

[0065] 在一个实施例中,一种或多种探针(例如化学探针)可与未修饰的胞嘧啶和5-甲基-修饰的胞嘧啶进行不同的反应。探针可选自于包括亚硫酸氢钠、焦亚硫酸钠和/或亚硫酸氢盐的一组。

[0066] 在另一个实施例中,分离的基因组DNA或核DNA与一种或多种探针反应的结果可以是未修饰胞嘧啶脱氨成尿嘧啶和未改变的5-甲基胞嘧啶。表征使分离的基因组DNA或核DNA与一种或多种探针探测(或反应)或分析甲基化状态结果可以涉及到DNA扩增和核酸序列的测定及检测与年龄相关的表观遗传标记中CpG二核苷酸C位或是胞嘧啶或是胸腺嘧啶的存在。进一步地,DNA扩增之后可能是噬菌体RNA聚合酶的转录、核糖核酸酶裂解和核糖核酸酶裂解产物的基质辅助激光解吸电离飞行时间质谱(MALDI-TOF MS)。

[0067] 核酸序列的测定可能涉及到一个或多个下列程序:核酸片段化、限制性内切酶消化、核酸杂交、引物延伸、焦磷酸测序、单核苷酸延伸、具有生物素标记的ddNTP的单核苷酸延伸、具有2,4-二硝基苯酚(DNP)标记的ddNTP的单核苷酸延伸、放射性同位素标记、非放射性标记掺入、荧光标记掺入、生物素掺入、抗原-抗体复合物的形成、抗体检测、比色检测、荧光检测、用荧光染料标记的抗体的检测、标记的抗生物素蛋白或抗生蛋白链菌素的检测、磁珠分析或检测方法、信号放大、聚合酶链反应、具有热稳定DNA聚合酶的DNA扩增、phi-29DNA聚合酶DNA扩增、RNA的合成、体外转录、噬菌体RNA聚合酶转录、T7 RNA聚合酶的转录、SP6RNA聚合酶的转录、T3RNA聚合酶转录、RNA酶消化、RNA酶A消化、DNA克隆、细菌转化、凝胶电泳、质谱法、MALDI-TOF质谱法、微阵列分析、荧光扫描分析、自动化的数字图像捕捉、自动化的数字图像分析、比率分析和 Infinium® HumanMethylation450磁珠芯片分析。

[0068] 在另一个实施例中,最初呈现在每个与年龄相关的表观遗传标记未修饰的胞嘧啶的比例可以为在CpG二核苷酸嘧啶位置具有胸腺嘧啶的与年龄相关的表观遗传标记的分数或者百分比。

[0069] 最初呈现在每个与年龄相关的表观遗传标记的5-甲基胞嘧啶和胞嘧啶的比可以是暴露于一种或多种探针后CpG二核苷酸嘧啶位置上的胞嘧啶与胸腺嘧啶的比,随后进行核酸扩增产品的分析。

[0070] 在另一个实施例中,最初呈现在每个与年龄相关的表观遗传标记中的未修饰的胞

嘧啶和5-甲基胞嘧啶的比可以是暴露于一种或多种探针后CpG二核苷酸嘧啶位置上的胸腺嘧啶与胞嘧啶的比,随后进行核酸扩增产品的分析。

[0071] 在一个实施例中,受到CpG二核苷酸的影响,在内部或接近限制性内切酶的鉴别点或者分裂位点,及基于在所述CpG二核苷酸的胞嘧啶的C-5位位置上存在或不存在甲基基团DNA裂解方式的不同,可以选择一种或多种可鉴别特定DNA序列的限制性内切酶。此类限制性内切酶的例子包括但不限于AatII、Acc65I、AccI、AciI、AcII、AfeI、AgeI、AhdI、AieI、ApaI、ApaLI、AscI、AsiSI、AvaI、AvaII、BaeI、BanI、BbrPI、BbvCI、BceAI、BcgI、BcoDI、BfuAI、BfuCI、BglI、BmgBI、BsaAI、BsaBI、BsaHI、BsaI、BseYI、BsiEI、BsiWI、BslI、BsmAI、BsmBI、BsmFI、BspDI、BsrBI、BsrFI、BssHII、BssKI、BstAPI、BstBI、BstUI、BstZ17I、Cac8I、ClaI、DpnI、DraIII、DrdI、EaeI、EagI、EarI、EciI、Eco53kI、EcoRI、EcoRV、FauI、Fnu4HI、FokI、FseI、FspI、HaeII、HaeIII、HgaI、HhaI、HincII、HinfI、HinPII、HpaI、HpaII、Hpy166II、Hpy188III、Hpy99I、HpyAV、HpyCH4IV、HpyCH4V、KasI、MboI、MluI、MmeI、MspAII、MwoI、NaeI、NarI、NciI、NgoMIV、NheI、NlaIV、NotI、NruI、Nt.BbvCI、Nt.BsmAI、Nt.CviPII、Pae71、PhoI、PleI、PfuTI、PmeI、PmlI、PshAI、PspOMI、PspXI、PvuI、RsaI、RsrII、SacII、SaiI、Sau3AI、Sau96I、ScrFI、SfaNI、SfiI、SfoI、SgrAI、SmaI、SnaBI、StyD4I、TfiI、TliI、TseI、TspMI、XhoI、XmaI,和ZraI。

[0072] 在一个实施例中,分离的基因组DNA或核DNA与一种或多种限制性内切酶探针反应的结果可以是当CpG二核苷酸中的胞嘧啶未修饰时在限制性内切酶分裂位点产生一个双链DNA断裂或者是当CpG二核苷酸中的胞嘧啶在其C-5位用甲基基团修饰时在限制性内切酶分裂位点无DNA双链断裂。

[0073] 在另一个实施例中,分析甲基化状态可以包括DNA扩增和由于限制性内切酶的分裂而特定DNA端(多个)或片段(多个)及与特定年龄相关的表观遗传标记相关联的完整限制性内切酶分裂位点的与年龄相关的表观遗传标记的分析。

[0074] 在另一个实施例中,最初呈现在每个与年龄相关的表观遗传标记未修饰的5-甲基胞嘧啶的比例可以为由于限制性内切酶的耐分裂具有完整限制性内切酶分裂位点与年龄相关的表观遗传标记的分数或者百分比。

[0075] 在另一个实施例中,最初呈现在每个与年龄相关的表观遗传标记未修饰的胞嘧啶的比例可以为限制性内切酶的分裂的与年龄相关的表观遗传标记的分数或者百分比。

[0076] 在另一个实施例中,最初呈现在每个与年龄相关的表观遗传标记的5-甲基胞嘧啶与未修饰的胞嘧啶的比可以为完整地限制性内切酶分裂位点的数目或浓度与由与年龄相关的表观遗传标记的限制性内切酶所产生的DNA双链断裂数目或浓度之比。

[0077] 在另一个实施例中,最初呈现在每个与年龄相关的表观遗传标记的未修饰的胞嘧啶与5-甲基胞嘧啶的比可以为由限制性内切酶所产生的DNA双链断裂的数目或浓度与由与年龄相关的表观遗传标记的5-甲基胞嘧啶的存在抗限制性内切酶分裂的限制性内切酶分裂位点的数目或浓度之比。

[0078] 在一个实施例中,确定甲基化状态包括从样品中分离基因组DNA或核DNA。另外,该步骤涉及用一种或多种探针探测分离的基因组DNA或核DNA及扩增DNA,所述的一种或多种探针可与未修饰的胞嘧啶和5-甲基-修饰的胞嘧啶进行不同的反应。该步骤还涉及用一种或多种限制性内切酶消化扩增的DNA,所述的一种或多种限制性内切酶鉴别含有CpG二核苷

酸的限制性内切酶位点,但不能消化从CpG二核苷酸突变为TpG二核苷酸的限制性内切酶位点。进一步地,该步骤涉及基于敏感或耐消化的限制性内切酶位点的分数或百分比确定最初呈现在每个与年龄相关的表观遗传标记的5-甲基胞嘧啶或胞嘧啶的比例。

[0079] 另外,甲基化状态可以涉及基于敏感性消化限制性内切酶位点的数目或浓度与耐消化限制性内切酶位点的数目或浓度之比,确定最初呈现在每个与年龄相关的表观遗传标记的5-甲基胞嘧啶与未修饰的胞嘧啶的比。甲基化状态还可以涉及基于敏感性消化限制性内切酶位点的数目或浓度与耐消化限制性内切酶位点的数目或浓度之比,确定最初呈现在每个与年龄相关的表观遗传标记的未修饰的胞嘧啶与5-甲基胞嘧啶的比。

[0080] 在另一个实施例中,确定选自图9、表S3、S4和/或S5受试者基因组的与年龄相关的一套表观遗传标记甲基化状态可以包括分离 基因组DNA或核DNA及破碎基因组DNA或核DNA。另外,所述步骤涉及把破碎的DNA暴露于5-甲基胞嘧啶结合蛋白中,及从5-甲基胞嘧啶结合蛋白-自由DNA片段中分离5-甲基胞嘧啶结合蛋白结合的DNA片段。该步骤进一步涉及通过确定每个与年龄相关的表观遗传标记的5-甲基胞嘧啶结合蛋白结合或游离的DNA片段的分数或百分比,确定每个与年龄相关的表观遗传标记的含DNA片段的5-甲基胞嘧啶或含DNA片段的未修饰的胞嘧啶的比例。

[0081] 此外,甲基化状态可以涉及通过确定5-甲基胞嘧啶结合蛋白结合的DNA片段的数目或浓度与5-甲基胞嘧啶结合蛋白-自由DNA片段的数目或浓度的比,确定每个与年龄相关的表观遗传标记的含DNA片段的5-甲基胞嘧啶与含DNA片段的未修饰的胞嘧啶的比例。

[0082] 甲基化状态步骤还可以涉及通过确定5-甲基胞嘧啶结合蛋白-自由DNA片段的数目或浓度与5-甲基胞嘧啶结合蛋白结合的DNA片段的数目或浓度的比,确定每个与年龄相关的表观遗传标记的含DNA片段的未修饰的胞嘧啶与含DNA片段的5-甲基胞嘧啶的比例。

[0083] 在一个实施例中,5-甲基胞嘧啶结合蛋白可以是5-甲基胞嘧啶、MeCP2、MBD2、MBD2/MBD3L1复合物、MBD2的核MBD区域、或者聚MBD蛋白、天然存在的5-甲基胞嘧啶结合蛋白、遗传工程5-甲基胞嘧啶结合蛋白、或其衍生物或其片段的抗体。

[0084] 在一个实施例中,从5-甲基胞嘧啶结合蛋白-自由DNA片段中分离5-甲基胞嘧啶结合蛋白-结合DNA片段可以包括免疫沉淀法、免疫捕获法、固相色谱法、液相色谱法和/或凝胶电泳法。

[0085] 按照本发明的实际,本发明提供了用于确定受试者表观甲基化衰老速率(AMAR)的方法。该方法包括通过本发明的方法预测年龄及把预测的年龄除以实际实足年龄。

[0086] 在一个实施例中,本发明提供了用于诊断受试者肿瘤存在的方法。该方法包括获得疑似含有肿瘤的生物样品及已知不含有肿瘤的第二份生物样品。此外,该方法包括通过本发明的方法预测各个生物样品的年龄。该方法还包括对比两种样品所预测的年龄,得到具有肿瘤的生物样品比无肿瘤的生物样品预测年龄较年长。

[0087] 在一个实施例中,本发明提供了一种从人类组织通过基于本发明方法的表观遗传标记预测受试者的年龄的人类实际年龄的法医诊断。

[0088] 在另一个实施例中,本发明提供了通过基于本发明方法的受试者表观遗传组预测受试者的年龄的受试者健康评估的方法。

[0089] 在另一个实施例中,本发明提供了筛选有益试剂是延缓或加速衰老过程的方法。该方法包括从活有机体中获得生物样品,及所述的有机体优选为从活有机体衍生的培养细

胞、组织或器官,及预测年龄或有机体的AMAR,随本发明的方法之后,使用适当的与年龄相关的表观遗传标记的有机体,以便人类受试者与年龄相关的表观遗传标记可以需要被已检查过的有机体的与年龄相关的表观遗传标记所取代。另外,该方法包括把有机体或来自活有机体的培养活细胞、组织或器官暴露于单剂量、多剂量或连续剂量的有益试剂中,从而从活有机体或培养活细胞、组织或器官中获得生物样品。该方法还包括使用适当的与年龄相关的表观遗传标记的有机体预测年龄或生物样品的有机体的AMAR,以便人类受试者的与年龄相关的表观遗传标记可以需要被已检查过的有机体的与年龄相关的表观遗传标记所取代。该方法还包括在另一个个体上实施相同的步骤,该个体来自于同样的生物或完全相同的培养的活细胞、组织或者同一个体的器官或者未用有益试剂处理或用安慰剂处理过的有机体,及比较相同预测年龄或AMAR的生物样品,预测年龄或用有益试剂处理过的有机体/个体、培养细胞、组织的AMAR与器官的预测年龄或未处理或用安慰剂处理过的有机体/个体或培养细胞、组织或器官的AMAR,得到用有益试剂处理过的有机体/个体或培养细胞、组织或器官的一个较低的值或值的范围表明有益试剂可减缓衰老过程而较高的值或值的范围表明有益试剂可加速衰老过程。有益试剂可以是抗氧化剂、还原剂、DNA损伤剂、维生素、营养补充剂、食品、食品添加剂、食品着色剂、盐、蔬菜、蔬菜提取物、水果、水果提取物、花、花提取物、香料、种子、籽提取物、香草、草本提取物、植物提取物、纤维、脂肪、脂肪酸、油、糖、人造甜味剂、益生菌、乙醇、葡萄酒、真菌、霉菌、奶油、洗剂、粉末、化妆品、防晒剂、气体、污染物、烟雾、环境污染物、油漆、溶剂、有机溶剂、塑料、增塑剂、双酚、酚类化合物、烟草、吸入剂、药物、生物制品、激素、内分泌干扰物、环境雌激素、激素拮抗剂、激素激动剂,咖啡因、植物雌激素、金属、酶、螯合剂、酸奶、硫化物,物理屏障、电磁障碍和辐射屏障。

[0090] 在一个实施例中,有机体可以是酵母、果蝇、鱼类、蠕虫、昆虫、斑马鱼、线虫、植物或哺乳动物。哺乳动物包括但不限于为人类、鼠、猴、猫、犬、马、牛、猪、绵羊、山羊、兔、哺乳类家畜、哺乳类运动动物和哺乳类宠物。

[0091] 在一个实施例中,本发明提供了用于鉴别已知年龄受试者的生物样品的组织类型的方法。该方法包括确定受试者的实足年龄及通过本发明方法确定来自于生物样品的受试者的预测年龄。另外,本方法包括对比各种组织类型的相关预测年龄的参考标准与实足年龄,及确定各种组织类型参考标准中与预测年龄最匹配的值。进一步地,该方法包括基于最匹配指定生物样品的组织类型。

[0092] 本发明还提供了一种用于鉴别已知实足年龄受试者的生物样品的组织类型的方法。该方法包括确定受试者的实足年龄及通过从受试者实足年龄分出受试者预测年龄,确定来自于生物样品的受试者的AMAR。另外,该方法包括对比相关AMAR的参考标准与各种组织类型的实足年龄,及确定各种组织类型参考标准的AMAR最匹配的值。该方法进一步包括基于最匹配指定生物样品的组织类型。

[0093] 在一个实施例中,与年龄相关的一套标记包括任何一种或多种图9、表S3和S5的cg23606718和/或图9、表S3和S5的cg16867657表观遗传标记。

[0094] 本发明进一步提供了基于表观遗传修饰影响基因表达预测受试者年龄的方法。该方法包括获得受试者的生物样品及确定与年龄相关的表观遗传标记相关联的其表达随着年龄变化的一个或多个的基因的表达。另外,该方法包括对比与年龄相关的表观遗传标记相关联的其表达随着年龄变化的一个或多个的基因的表达与来自与年龄相关的参考人群

的相同基因的表达。该方法进一步包括获得受试者预测年龄的一个值或值的范围。对比与年龄相关的表观遗传标记相关联的其表达随着年龄变化的一个或多个的基因的表达与来自与年龄相关的参考人群的相同基因的表达可以包括使用任一统计方法、多变量回归分析法、线性回归分析、表格法、或图解法基于与年龄相关的表观遗传标记相关联的其表达随着年龄变化的基因的表达预测受试者的年龄。在一个实施例中，统计方法可以是多元回归法或线性回归法。

[0095] 在另一个实施例中，与年龄相关的表观遗传标记相关联的其表达随着年龄变化的一个或多个的基因的表达可以包括一个或多个列在表S6或表S7的基因。

[0096] 在另一个实施例中，基因表达可以是转录或翻译。在另一个实施例中，转录导致RNA转录体的产生，翻译导致蛋白质的产生。

[0097] 按照本发明的实际，本发明提供了一种用于筛选受试者组织样品以基于由本发明方法的受试者的表观基因组预测组织样品年龄的方法。

[0098] 在一个实施例中，在高通量的筛选试验中组织样品可以暴露于至少一种测试试剂。在另一个实施例中，所述过程可以用于任一诊断和/或高通量筛选。

[0099] 本发明进一步提供了基于受试者的组织或器官的表观基因组预测受试者的组织或器官年龄的方法。该方法包括从受试者获得组织或器官的生物样品，及在选自于图9、表S3、S4和/或S5受试者表观基因组中确定与年龄相关的一套表观遗传标记的甲基化状态。该方法进一步包括对比受试者与年龄相关的一套表观遗传标记的甲基化状态与来自与年龄相关的参考人群相同标记的甲基化状态，以获得预测组织或器官年龄的一个值或值的范围。

[0100] 来自与年龄相关的参考人群的相同标记的甲基化状态可以在相同或不同类型的组织或器官中确定。来自与年龄相关的参考人群相同标记的甲基化状态可以在血液或分离血液中确定。

[0101] 在一个实施例中，本发明的方法提供了用于确定受试者组织或器官不同衰老速率的方法。该方法包括从受试者不同组织或器官中获得生物样品及使用本发明的方法预测组织或器官的年龄。该方法进一步包括对比预测的年龄，其中预测年龄的差别表明受试者组织或器官衰老速率的差别。预测年龄可除以受试者的实足年龄以获得AMAR。

[0102] 本发明的组合物

[0103] 本发明还提供了组合物，所述组合物包括一套表观遗传标记，所述表观遗传标记基于选自图9，表S3、S4和/或S5受试者表观基因组的5个或5个以上的与年龄相关的表观遗传标记；选自图9，表S3、S4和/或S5受试者表观基因组的10个或10个以上的与年龄相关的表观遗传标记；选自图9，表S3、S4和/或S5受试者表观基因组的15个或15个以上的与年龄相关的表观遗传标记；选自图9，表S3、S4和/或S5受试者表观基因组的20个或20个以上的与年龄相关的表观遗传标记；选自图9，表S3、S4和/或S5受试者表观基因组的25个或25个以上的与年龄相关的表观遗传标记；选自图9，表S3、S4和/或S5受试者表观基因组的30个或30个以上的与年龄相关的表观遗传标记；选自图9，表S3、S4和/或S5受试者表观基因组的35个或35个以上的与年龄相关的表观遗传标记；选自图9，表S3、S4和/或S5受试者表观基因组的40个或40个以上的与年龄相关的表观遗传标记；选自图9，表S3、S4和/或S5受试者表观基因组的45个或45个以上的与年龄相关的表观遗传标记；选自图9，表S3、S4和/或S5受试者表观基因

组的50个或50个以上的与年龄相关的表观遗传标记;选自图9,表S3、S4和/或S5受试者表观基因组的55个或55个以上的与年龄相关的表观遗传标记;选自图9,表S3、S4和/或S5受试者表观基因组的60个或60个以上的与年龄相关的表观遗传标记;选自图9,表S3、S4和/或S5受试者表观基因组的65个或65个以上的与年龄相关的表观遗传标记;选自图9,表S3、S4和/或S5受试者表观基因组的70个或70个以上的与年龄相关的表观遗传标记。

[0104] 进一步地,在优选的实施例中,所述组合物包括一套表观遗传标记,所述表观遗传标记基于图9或表S3受试者表观基因组的5个或5个以上的与年龄相关的表观遗传标记;图9或表S3受试者表观基因组的10个或10个以上的与年龄相关的表观遗传标记;图9或表S3受试者表观基因组的15个或15个以上的与年龄相关的表观遗传标记;图9或表S3受试者表观基因组的20个或20个以上的与年龄相关的表观遗传标记;图9或表S3受试者表观基因组的25个或25个以上的与年龄相关的表观遗传标记;图9或表S3受试者表观基因组的30个或30个以上的与年龄相关的表观遗传标记;图9或表S3受试者表观基因组的35个或35个以上的与年龄相关的表观遗传标记;图9或表S3受试者表观基因组的40个或40个以上的与年龄相关的表观遗传标记;图9或表S3受试者表观基因组的45个或45个以上的与年龄相关的表观遗传标记;图9或表S3受试者表观基因组的50个或50个以上的与年龄相关的表观遗传标记;图9或表S3受试者表观基因组的55个或55个以上的与年龄相关的表观遗传标记;图9或表S3受试者表观基因组的60个或60个以上的与年龄相关的表观遗传标记;图9或表S3受试者表观基因组的65个或65个以上的与年龄相关的表观遗传标记;图9或表S3受试者表观基因组的70个或70个以上的与年龄相关的表观遗传标记。例如,与年龄相关的一套表观遗传标记可包括任何一个或多个下列的甲基化标记:表S4的cg05652533、表S4的cg27367526、表S4的cg18404041、图9、表S3和S5的cg23606718、图9、表S3和S5的cg16867657、位置52008487上染色体3的cg04474832、位置21369010上染色体22的cg05442902、位置25652602上染色体6的cg06493994、位置236557682上染色体1的cg09809672、位置30849114上染色体12的cg19722847、位置18122719上染色体6的cg22736354、表S4的cg05652533、表S4的cg27367526、表S4的cg18404041、位置131513927上染色体2的cg23606718和/或位置11044877上染色体6的cg16867657。

[0105] 在另一个实施例中,所述组合物包括与年龄相关的一套表观遗传标记:位置上52008487染色体3的cg04474832、位置21369010上染色体22的cg05442902、位置25652602上染色体6的cg06493994、位置236557682上染色体1的cg09809672、位置30849114上染色体12的cg19722847和位置18122719上染色体6的cg22736354。

[0106] 在另一个实施例中,所述组合物包括任何一个或多个甲基化标记的与年龄相关的一套表观遗传标记,所述的任何一个或多个甲基化标记为:图9或表S3的cg20822990、图9或表S3的cg04400972、图9或表S3的cg16054275、图9或表S3的cg03607117、图9或表S3的cg20052760、图9或表S3的cg16867657、图9或表S3的cg06493994、图9或表S3的cg06685111、图9或表S3的cg00486113、图9或表S3的cg20426994、图9或表S3的cg14361627、图9或表S3的cg08097417、图9或表S3的cg07955995、图9或表S3的cg22285878和/或图9或表S3的cg08540945。

[0107] 在另一个实施例中,所述组合物包括任何一个或多个甲基化标记的与年龄相关的一套表观遗传标记,所述的任何一个或多个甲基化标记为:图9、表S3和S5的cg23606718和/

或图9、表S3和S5的cg16867657。

[0108] 试剂盒

[0109] 根据本发明的另一方面,提供了试剂盒。依据本发明的试剂盒包括装有本发明化合物或组合物的包装盒。

[0110] 术语“包装盒”是指装有本文所出现的化合物或组合物的任何容器。在优选的实施例中,包装盒可为盒子或包装纸。用于包装药物产品的包装材料是本领域技术人员公知的技术。药物包装材料的例子包括但不限于罩板包装、瓶、试管、吸入器、泵、袋、小瓶、容器、注射器、瓶和适于选择配方及给药和治疗的预期模式的任何包装材料。

[0111] 该试剂盒也可包括未装在包装盒内而连接到包装盒外面的物品,如吸量管。

[0112] 该试剂盒可任选包括用于将本发明的化合物或组合物给药于具有需要治疗条件的受试者的说明书。该试剂盒还可包括监管机构如美国食品和药物管理局在此认可化合物使用的说明书。该试剂盒可任选包括该化合物的标签或产品插页。该包装盒和/或任何产品插页本身可通过监管机构认可。该试剂盒可包括包装盒中固相或液相(如提供的缓冲液)的化合物。该试剂盒也可包括用于制备溶液的缓冲液以实施该方法,还及用于把液体从一个容器转移到另一个容器的吸量管。

[0113] 该试剂盒可任选包括描述于本文中组合治疗法所用的一种或多种其他化合物。在某个实施例中,静脉注射中该包装盒是一个容器。在其他实施例中,用吸入器提供化合物。仍然在其他实施例中,用聚合物基质或脂质体的形式提供化合物。

[0114] 本发明提供了一种基于受试者遗传物质的表观遗传修饰确定受试者年龄的试剂盒,包括与年龄相关的一套表观遗传标记或在本发明方法描述的图9、表S3、表S4或表S5所列举的标记,同上。

[0115] 本发明还提供了一种基于受试者的表观遗传标记确定受试者年龄的试剂盒,使用在本发明方法描述的图9、表S3、S4和/或S5所提供的与年龄相关的一套表观遗传标记,同上。

[0116] 在一个实施例中,与年龄相关的表观遗传标记可包括具有CpG二核苷酸的核酸。在另一个实施例中,所述CpG二核苷酸的胞嘧啶在C-5位的甲基化受年龄依赖性变化的影响。在另一个实施例中,所述CpG二核苷酸在如图9、表S3、S4和/或S5所描述的染色体位置。

[0117] 在一个实施例中,与年龄相关的表观遗传标记可为人类的标记,并选自于位置52008487上染色体3的cg04474832、位置21369010上染色体22的cg05442902、位置25652602上染色体6的cg06493994、位置236557682上染色体1的cg09809672、位置30849114上染色体12的cg19722847、位置18122719上染色体6的cg22736354、表S4的cg05652533、表S4的cg27367526、表S4的cg18404041、位置131513927上染色体2的cg23606718和位置11044877上染色体6的cg16867657。

[0118] 在另一个实施例中,与年龄相关的表观遗传标记可具有如图9所提供的序列或者是在国家卫生研究院生物技术信息国家中心(Bethesda, MD)具有GEO登录号为GPL13534的基因表达综合(GEO)数据库中可找到的序列。

[0119] 提供以下实施例进一步解释本发明的方面,这些实施例是非限制的,并且不应该被解释为限制本发明的任何方面。

实施例

[0120] 实施例1

[0121] 材料和方法

[0122] 样品采集和测试程序

[0123] 该研究得到圣地亚哥的美国加州大学的机构审查委员会、南加州大学和华西医院的认可。所有与会者在参与前都签署知情同意书声明。血液从病人手臂的静脉抽取进入含有抗凝血酸柠檬酸葡萄糖的血液收集管中。基因组DNA用在 -20°C 下存放的QIAGEN FLEXiGene DNA试剂盒从全血中抽取并。常染色体的甲基化分值用Illumina Infinium HumanMethylation450芯片(Bibikova et al., 2011)测定。本程序对可结合到甲基化和未甲基化的的序列中的每个标记使用硫酸氢盐处理的DNA和两位点特异性探针,所述每个标记。甲基化探针强度相对于每个位点的总探针强度代表样品中该位点甲基化的分数水平。这些值是用Illumina公司基因组办公软件进行内部控制校对的。具有高于0.01的检测P值的甲基化分值被设置为“缺失(missing)”。由于一个病人样品及830个标记具有高于5%的缺失值,因而被除去。余下的缺失值使用R“输入”包(Troyanskaya et al., 2001)通过KNN法(10个最接近的标记)输入。我们在这些样品258上进行外显子组测序,使用溶液混合选择方法以捕获DNA,随后在Illumina HiSeq平台上进行平行测序。基因型调用是由SOAP程序作出(Li et al., 2008)。质量分数低于20的调用被设置成缺失。只有不到10%缺失调用的变量在哈迪-温伯格(Hardy-Weinberg)平衡($p < 10^{-4}$)内,及为共同频率($> 5\%$)保留下来(10, 694)。不到20%缺失调用(252)的个体被保留下来。另外的基因型用多重PCR处理,随后用带有iPLEX/MassARRAY/Typer平台的MALDI-TOF质谱分析。

[0124] 甲基化的质量控制

[0125] 我们使用主成分(PC)分析来鉴别和除去异常样品。基于从总体平均数得来的样品第一个PC的平方距离,我们把每一个样品转换成 Z得分统计量。用高斯累积分布和Benjamini-Hochberg程序(Benjamini-Hochberg, 1995)把Z统计量转化为错误发现率。落在低于0.2的FDR的样品下降被指定为异常并被除去。反复进行这种过滤程序直到无样品被确定为异常样品。以这种方式总共24个样品被除去。

[0126] 关联测试

[0127] 甲基化分数和偏差中趋势的关联测试用嵌套线性模型和F测试在进行。由于甲基化水平对许多因素敏感,我们列入几个协变量,包括性别、BMI、糖尿病病态、种族和批次。基于带有正系数而不是负系数的标记数目,整个甲基化异常变化的试验用二项检验法计算。当衰老系数为相同的符号及显著性好于 $p < 0.05$ 时标记被注释为具有TCGA数据的支撑,。

[0128] 注释富集

[0129] 从来自于Bioconductor(Gentleman et al., 2004)的IlluminaHumanMethylation450k.db数据库中获得CpG岛的甲基化标记注释和GO术语。注释富集测试用双面Fisher精确测试进行。

[0130] 衰老模型

[0131] 基于在R包装盒“glmnet”中实施弹力网络算法(Friedman et al., 2010),用多元线性模型方法进行年龄诊断模型。该方法为传统的套索工具与岭回归分析方法的结合,强

调模型稀疏,同时适当平衡相关变量的贡献。在变量的数量(标记)远远超过样品的数量的情况下构建线性模型是理想的。优化正则化参数通过10倍交叉验证评估。我们采用自展分析,采样数据设置更换500次,为每个引导队列建立一个模型。在最终的模型中,我们仅列入其出现占有所有引导程序的一半以上的标记。性别协变量、BMI、糖尿病病态、种族和批次列入模型且从零对偶间隙(正则化)中免除。基于最小二乘模型用相同的术语和drop-oneF测试建立P值。由于BMI与年龄密切相关,在计算模型中显著性以前,为年龄第一次调整术语。用衰老模型但不用性别、体重指数和糖尿病状态的变量计算AMAR。该系数是不变的。然后,AMAR被作为个体的预测年龄除以她或他的实际年龄。

[0132] 基因方差关联

[0133] 以具有与年龄相关的甲基化标记的加法模型用嵌套的线性模型和F测试测试每个基因方差的关联性。我们列入性别、BMI、糖尿病状态和批次的协变量。变异位置以构建GRCh37的人类参考为基础,基因注释以染色体接近于20kbp为基础。

[0134] 计算甲基化异常

[0135] 甲基化异常通过下述方法计算:首先,由于所有给定的包括年龄、性别和BMI的变量,通过拟合每个标记的线性模型,并且只在残留物上作用,我们去甲基化的趋势。其次,在夏皮罗-威尔克(Shapiro-Wilk)测试($p < 10^{-5}$)的基础上,我们鉴别并除去高度非正常标记。在正态性测试中,允许自然发生的极端异常,首先,基于格拉布(Grubb)的统计,我们估计每个标记的异常值,基于Benjamini-Hochberg FDR(Benjamini and Hochberg, 1995),选择包含阈值。如果任何样品的FDR低于0.4,我们忽视他们并重复异常值的检测,直到无异常值检出。最后,每个剩余标记的异常值被计算为其调整后甲基化值的平方。

[0136] 熵分析

[0137] 熵统计量进行了甲基化数据计算,协变量调整,并过滤成标准值(参见计算甲基化异常值)。我们根据公式计算每个个体甲基化的标准Shannon熵(Shannon and Weaver, 1963)。

[0138]

$$Entropy = \frac{1}{N * \log(\frac{1}{2})} \sum_i [MFi * \log(MFi) + (1 - MFi) * \log(1 - MFi)]$$

[0139] 其中,MFi是i次甲基化标记的甲基化分值,及N为标记的数目。

[0140] 测绘CpG岛

[0141] 从来自于Bioconductor(Gentleman et al., 2004)的IlluminaHumanMethylation450k.db数据库中获得27,176CpG岛的基因位置和标记注释。我们获得具有至少四个标记(25,028)的每一个岛的标记的位置及最近的100个上游和下游标记。然后,这些位置与有益的标记值(如甲基化分数、衰老系数或异常值)结合以产生每个岛及周边地区的基因组地图。在把每一副地图规范到岛的中心后,我们把所有岛的每个基因组的相对点的值平均以产生一个通用地图。

[0142] 结果和讨论

[0143] 宽年龄范围内的全局甲基化图谱

[0144] 从426个白种人和236个西班牙个体的混合人群中,年龄从19到101,我们获得两个不同群体(N1=482,N2=174)的甲基化谱宽图谱。样品被作为全血及用测定485,577CpG标

记的甲基化状态的Illumina Infinium HumanMethylation450磁珠芯片分析仪(Bibikova et al., 2011)处理。甲基化被记录为0至1之间的分数,代表来自于在人群中单一个体血细胞所给的CpG标记的甲基化频率。传统的质量控制应用于过滤假标记和样品。为了简化,我们舍弃性染色体标记上的值。相关性测试显示出70387(15%)的标记在甲基化分值和年龄之间具有显著的相关性(图1,通过F测试错误发现率[FDR]<0.05)。我们能确认海恩等人(Heyn et al.)最近公开的(2012)在 $p < 0.05$ 显著水平上用40个年轻和年老的样品的相关性为53,670(76%)。图S1和表S2和S3给出了个体衰老标记及其基因特性的更详细的分析。所得到的数据集代表了由衰老研究产生的规模最大、分辨率最高的甲基化数据集合,在衰老过程中了解表观遗传学的作用提供了一个前所未有的机会。在基因表达数据库中可得到完整的甲基化图谱(GSE40279)。

[0145] 表S1,与图1有关的与年龄相关标记的功能注释

[0146] 与年龄相关的标记附近的基因富集了许多功能,精选的功能如下所示。

[0147]

超富集术语	低富集术语
细胞通讯 (FDR~ 10^{-76})	G-蛋白偶联受体活性 (FDR < 10^{-11})
移动 (FDR< 10^{-33})	核糖体 (FDR< 10^{-5})
细胞增殖 (FDR< 10^{-17})	RNA 剪接 (FDR<0.05)
生长 (FDR< 10^{-7})	有丝分裂期 (FDR<0.05)

[0148] 表S2,与图1有关的与年龄相关的标记属性

[0149] 一个与年龄相关的标记及其与几个基因组特征的一致性的表。每一个值代表了特定类型(列)的与年龄相关的标记的百分比,所述的特定类型与特定的注释(行)是一致的。

[0150]

	所有标记 (%)	递增均值 (%)	递减均值 (%)	递增方差 (%)
MF>0.5	55	13	57	51
MF<0.5	45	87	43	49
CpG 岛	31	68	8	24
CpG 岸	23	18	35	36
CpG 架	10	3	13	8
增强剂	22	22	30	27
启动子	20	7	13	8
DHS	12	30	14	16

[0151] 衰老甲基化的预测模型

[0152] 用被称为弹性网络的惩罚多元回归方法(Zou and Hastie,2005),结合自举法,我们在主要行列构建了衰老的预测模型。该模型包括甲基化和如性别和体重指数(BMI)(图2A)的临床指标。优化模型选自于具有年龄高度预测性的一套71个甲基化标记(图2A和表S3)。模型的精度高,年龄与预测年龄之间的相关性为96%及误差3.9年(图2B)。在模型中几乎所有的标记放置在与衰老相关的条件下已知功能基因内或附近,与衰老相关的条件包括阿尔茨海默氏病、癌症、组织降解、DNA损伤和氧化应激。以举例的方式,两个标记放置在内分泌和神经系统功能的一个重要调节器(Yacubova and Komuro,2002)的基因生长抑素(SST)内。众所周知,SST是随着年龄的增长而衰退,且与阿尔茨海默氏病相关联(Saito et al.,2005)。作为第二个例子,六个模型标记放置在转录因子KLF14内,所述的转录因子KLF14被称为肥胖和其他代谢特征的“主调节器”(SmaII et al.,2011)。考虑到衰老、长寿和新陈代谢活动之间的关联(Lane et al.,1996;Tatar et al.,2003),我们模型的几个标记与肥胖和代谢活动有关联是不足为奇的。

[0153] 表S3,与图2有关的衰老模型标记

[0154] 甲基化标记的一个表格包括在主要衰老模型中。每个标记所列举的系数是其模型中的回归系数。基于所有样品的模型(主要和验证)提供了第二个表格。

[0155]

标记	染色体	位置	基因组	CpG岛	系数
cg20822990	1	17338766	ATP13A2,SDHB	No	-15.7
cg22512670	1	26855765	RPS6KA1	No	1.05
cg25410668	1	28241577	RPA2,SMPDL3B	No	3.87
cg04400972	1	117665053	TRIM45,TTF2	Yes	9.62
cg16054275	1	169556022	F5,SELP	No	-11.1
cg10501210	1	207997020	Clorf132	No	-6.46
cg09809672	1	236557682	EDARADD	No	-0.74

ch.2.30415474F	2	30561970	LBH	No	5.79
cg22158769	2	39187539	ARHGEF33	Yes	-2.06
cg02085953	2	97202260	ARID5A	No	1.02
cg06639320	2	106015739	FHL2	Yes	8.95
cg22454769	2	106015767	FHL2	Yes	4.85
cg24079702	2	106015771	FHL2	Yes	2.48
cg23606718	2	131513927	FAM123C	Yes	8.35
cg22016779	2	230452311	DNER,RNU7-9P	No	1.79

[0156]

标记	染色体	位置	基因组	CpG 岛	系数
cg04474832	3	52008487	ABHD14A,ABHD14B, ACY1, GPR62,PCBP4,RPL29	No	-7.1
cg03607117	3	53080440	SFMBT1	Yes	10.7
cg07553761	3	160167977	SMC4,TRIM59	Yes	3.72
cg00481951	3	187387650	SST	No	-2.72
cg25478614	3	187387866	SST	No	4.01
cg25428494	4	84255411	HPSE	No	-1.81
cg02650266	4	147558239	POU4F2	Yes	10.2
cg08234504	5	139013317	CTD-3224K15.2,CXXC5, UBE2D2	No	-3.16
cg23500537	5	140419819	PCDHB1	No	5.67
cg20052760	6	10510789	GCNT2	No	-12.6
cg16867657	6	11044877	ELOVL2	Yes	10.8
cg22736354	6	18122719	NHLRC1,TPMT	Yes	4.42
cg06493994	6	25652602	SCGN	Yes	9.42
cg06685111	6	30295466	XXbac-BPG283O16.8	No	-13.1
cg00486113	6	31105711		No	-10.7
cg13001142	6	147528521	STXBP5	No	-5.8
cg20426994	7	130418324	KLF14	Yes	19.1
cg14361627	7	130419116	KLF14	Yes	10.7
cg08097417	7	130419133	KLF14	Yes	27.3
cg07955995	7	130419159	KLF14	Yes	13.7
cg22285878	7	130419173	KLF14	Yes	-20.7
cg03473532	7	131008743	MKLN1	No	-3.31
cg08540945	7	152591698		Yes	9.41
cg07927379	7	156433108	C7orf13,RNF32	Yes	-1.42
cg16419235	8	57360613	PENK	Yes	-1.6
cg07583137	8	82644012	CHMP4C,ZFAND1	No	3.03
cg22796704	10	49673534	ARHGAP22	No	-10.6
cg19935065	10	98062687	DNTT	No	13.4
cg23091758	11	9025767	NRIP3,SCUBE2	Yes	-0.392
cg23744638	11	10323902	ADM,AMPD3,SBF2	No	0.0859
cg04940570	11	12696758	TEAD1	Yes	11.6
cg11067179	11	66083541	CD248,RIN1,TMEM151A	No	14.7
cg22213242	11	66083573	CD248,RIN1,TMEM151A	Yes	23.7
cg06419846	11	66083697	CD248,RIN1,TMEM151A	Yes	13.4
cg02046143	11	133797911	IGSF9B	No	-10.2
cg00748589	12	11653486		Yes	8.21
cg19722847	12	30849114	CAPRN2,IPO8	No	-5.66
cg18473521	12	54448265	HOXC4,HOXC5	No	8.85
cg01528542	12	81468232	ACSS3	No	-2.98
ch.13.39564907R	13	40666907		No	-20.6

[0157]

标记	染色体	位置	基因组	CpG 岛	系数
cg03032497	14	61108227	SIX1	No	8.4
cg04875128	15	31775895	OTUD7A	Yes	-4.37
cg21296230	15	33010536	GREM1	Yes	8.39
cg09651136	15	72525012	PARP6,PKM2	No	-15.8
cg03399905	15	79576060	ANKRD34C	Yes	28
cg04416734	16	30075192	ALDOA,PPP4C	No	11.9
cg07082267	16	85429035		No	2.87
cg14692377	17	28562685	BLMH,SLC6A4,SNORD63.3	Yes	19.1
cg06874016	17	40177415	DNAJC7,NKIRAS2,ZNF385C	No	-4.37
cg21139312	17	55663225	MSI2	No	17.1
cg02867102	17	62398693		No	-12.5
cg19283806	18	66389420	CCDC102B,TMX3	No	-4.29
cg14556683	19	15342982	BRD4,EPHX3	Yes	-4.04
cg07547549	20	44658225	MMP9,SLC12A5	Yes	3.11
cg05442902	22	21369010	LZTR1,MIR649,P2RX6, SLC7A4,THAP7	No	-22.7
cg08415592	22	36648973	APOL1,APOL2,Z82215.1	No	-6.92
cg24724428					

[0158] 在第二个行列中我们验证了该模型,包括了额外的174个独立的样品。这些样品在主要行列中以同样的方式进行,然后,基于原有模型用于预测年龄(如在原有行列中培养过的)。预测高度准确,年龄与预测年龄之间的关联性为91%,及误差为4.9年(图2C)。通过海恩等人(Heyn et al.)所呈现的一组数据也确定了衰老模型的显著性,验证了71个标记中70个年龄的相关性(Heyn et al.,2012)。而且,在海恩等人(Heyn et al.)的研究中,该模型可以完全把年长个体和年轻个体分开,甚至通过亚硫酸盐测序获得图谱而不是使用本研究的磁珠芯片技术(图S2)。

[0159] 甲基化衰老速率及其相关性

[0160] 当衰老模型可以高准确度地预测大部分个体的年龄时,其相当于可以作为一种用于鉴别不遵循期望值的个别异常值的工具。例如,图2B强调了两个个体,其年龄以甲基化数据为基础大大高于或低于预测。为了检查这些差异是否在个体状态上反映了真正的生物学差异(如与测量误差或固有变异性相对),我们使用衰老模型来量化每个个体的显著的甲基化衰老速率(AMAR),基于甲基化数据,定义了预测年龄与实足年龄的比。然后,我们测试了AMAR和包括性别和BMI的可能相关的临床因素之间的关联性。由于批次变量的关联性(图S3),种族和糖尿病病态的分析是不可能的。我们发现,性别而不是BMI对衰老速率有显著地贡献(F测试, $p=6 \times 10^{-6}$, $p>0.05$)。即使在行列中年龄的总体分布在男人和女人之间的差异不明显($p>0.05$,KS测试),男人的甲基化比女人的甲基化衰老好像快约4%(图2D)。同样地,验证行列证实了男人的衰老速率的增加($p<0.05$),但对BMI是不确定的($p>0.05$)。这补充了随着年龄的变化BMI的表观遗传信号不发生变化的先前发现(Feinberg et al.,2010)。

[0161] 由于之前报道了基因相关性与人类长寿和基因表型的关系(Atzmon et al.,2006;Suh et al.,2008;Willcox et al.,2008;Wheeler et al.,2009),我们检查该模型

是否可区分具有不同基因变异的个体的衰老速率。为此,在15x覆盖范围内,在甲基化研究中,我们获得了252个个体的全外显子组序列。序列处理及质量控制后,这些序列在整个群体中产生了10,694个常见的单核苷酸变异体。作为阴性控制,我们证实了所有遗传变异都不能显著预测自己的年龄,这是可以预料的,因为基因组序列被认为在一生过程中是相对静止的。另一方面,人们可预期地发现调节与年龄相关的标记的甲基化的基因变异,如甲基化定量性状位点或meQTLs(BeII et al.,2011)。为测试每一个基因变异与顶部年龄相关联的甲基化标记的相关性,我们确定了303个meQTLs(FDR<0.05,图3A)。为了验证,我们选择了8个基因变异(对应于14 meQTLs)以测试来自于甲基化研究中332个在验证行列中的个体。本分析发现8个基因变异中(对应于7 meQTLs)的7个在验证行列体中仍保留高显著性(FDR<0.05,表S4)。而具有其meQTLs(在150 kbp内)的顺式起作用的所有变异,我们确认其没有一个可直接修饰CpG位点或者与相关甲基化标记的探针序列相关联。

[0162] 表S4. 与图2有关的基因变异影响与年龄相关的标记

[0163] 发现影响与年龄相关的甲基化的基因变异的一个表。间距为从基因标记到甲基化标记的基因组距离。相关值被列为P值。AMAR相关性为基因标记和AMAR之间的显著相关性。

[0164]

基因标记	遗传基因	甲基化标记	甲基化基因	间距 (bp)	甲基化-年龄相关性	甲基化-基因相关性	AMAR 相关性
rs2818384	JAKMIP3	cg05652533	DPYSL4, JAKMIP3	38793	5.86×10^{-9}	3.73×10^{-21}	0.00133
rs42663	GTPBP10	cg27367526	STEAP2	142116	1.44×10^{-18}	8.05×10^{-22}	0.00476
rs2230534	ITIH1, NEK4	cg18404041	ITIH1, ITIH3, NEK4	21881	6.78×10^{-14}	1.26×10^{-87}	0.02125
rs17152433	CTBP2, ZRANB1	cg07906193		70390	8.51×10^{-9}	7.54×10^{-56}	0.05273
rs1058405	ATF6	cg19735514	ATF6	10998	4.50×10^{-11}	3.87×10^{-63}	0.55546
rs57913893	ACSS1, C20orf3	cg26306437	C20orf3, CST7	59756	5.20×10^{-8}	5.50×10^{-16}	0.80327
rs6115003	ACSS1	cg26306437	C20orf3, CST7	70693	5.20×10^{-8}	4.17×10^{-17}	0.92096

[0165] 甲基化标记cg27193080是发现的与年龄有显著相关性的其中之一($p < 10^{-17}$),发现其甲基化分值受到单核苷酸多态性(SNP)变体rs140692的影响($p < 10^{-21}$)(图3B)。作为SNP和映射基因甲基-CpG结合结构域蛋白4(MBD4,具有基因内区SNP和仅在编码区的上游的甲基化标记)的甲基化标记的meQTL尤其引人关注,是少数已知基因编码可结合到甲基化DNA的蛋白质中的一个。因此,该meQTL可捕获顺式关系,其中rs140692影响MBD4的甲基化状态。该MBD4在人体衰老中的作用是支持把MBD4和DNA联系在一起的前期工作,及研究表明MBD4的突变及抑制导致的基因组不稳定性的增加,(BeIIacosa et al.,1999;Bertoni et al.,2009)。

[0166] 在7个已验证的meQTL中,确认了三个,不仅在年龄上而且在衰老速率上具有统计上的显著的相关性(AMAR,FDR<0.05,图3B和3C)。一个为基因标记rs2230534,在基因NEK4中是一个同义突变,并与甲基化标记cg18404041有顺式关联,NEK激酶家族在细胞周期调控和癌症中起着关键的作用(Moniz et al.,2011);第二变异为基因标记rs2818384,在基因

JAKMIP3中是一个同义突变并与甲基化标记cg05652533有顺式关联,JAKMIP3中复制数目的变异在先与成胶质细胞瘤相关联(Xiong et al.,2010);发现影响AMAR的最后一个变异为rs42663,在基因GTPBP10中是一个错义突变,并与STEAP2基因中的cg27367526相关联,众所周知,STEAP2在保持作为呼吸链的基本组成部分的金属-铁和铜的稳态中发挥着作用(Ohgami et al.,2006)。研究表明在哺乳动物细胞中通过氧化应激铁浓度的扰动可诱导DNA损伤(Hartwig and SchlepegreII,1995;Karthikeyan et al.,2002)。这些meQTL代表了基因变异,所述基因变异好像广泛影响了衰老甲基化,且可能是进一步研究年龄相关的疾病和长寿的好的候选者。

[0167] 多组织的诊断

[0168] 我们的衰老模型来自全血,其对实际诊断中的设计和从其他研究收集的测试样品是有利的。为了研究我们的衰老模型是否代表其他组织,我们从癌症基因组图谱(TCGA)控制类别的368个个体中获得DNA甲基化图谱(Collins and Barker,2007),包括83个胸脯、183个肾、60个肺和42个皮肤样品。尽管每个组织与预期都有明确的线性偏移量(截距和斜率)(图4A),基于我们主要和验证的行列的衰老模型对这些样品的实足年龄表现出强的预测能力具有(预期值 $R=0.72$)。这个偏移量在组织内是一致的,甚至横跨不同批次的TCGA数据。我们使用简单的线性模型调整每个组织的趋势,产生相对于血中所发现的具有误差可比性的年龄预测(图4B)。而且,在每个组织中所预测的AMAR支持了男人比女人衰老更快的效果($p<0.05$)。因此,从血样品中计算衰老速率(AMAR)反映了不特定于血液及可能在人体中许多组织中是常见的趋势。而且,这种分析提供了证据,所述证据为所观察到的甲基化变化是内在的甲基化,且主要不是由于细胞的异质性,如随着年龄的增长改变全血细胞类型的组成。在这一点上,本研究与纯化的CD4+T细胞和CD14+单核细胞的先前分析结果相一致,其中,发现与年龄相关的表观遗传修饰和在全血中所观察到的变化相类似(Rakyan et al.,2010)。

[0169] 为了研究组织间的相似性和差异性,我们重新构建了关于胸腺、肾和肺组织的衰老模型(表S5;皮肤行列有太少样品以致于不能构建模型)。尽管所有的模型和主要模型共享标记cg23606718和cg16867657,模型中的大多数标记不同。这些标记都被注释为基因ELOVL2,其与人体皮肤中的光衰老反应相联系(Kim et al.,2010)。

[0170] 表格S5.TCGA数据的衰老模型标记

[0171] 为了研究组织间的相似性和差异性,我们构建了关于胸腺、肾和肺组织的衰老模型。皮肤行列没有足够的样品来构建模型。每个模型的标记和系数列举如下:

[0172]

标记	染色体	位置	基因	CpG 岛	系数
cg23040782	1	6762215	DNAJC11	No	-7.45
cg11197101	1	33219998	KIAA1522	Yes	8.73
cg00252781	1	179334658	C1orf125,SOAT1	No	13.2
cg16909962	1	229406711	RAB4A,TMEM78	Yes	27.6
cg23606718	2	131513927	FAM123C	Yes	25.1
cg03545227	2	220173100	MIR153-1,PTPRN,RESP18	Yes	13.5
cg00702638	3	44803293	KIAA1143,KIF15	Yes	53.1
cg05555455	3	148804550	HLTF,Y_RNA.240	Yes	-18.5
cg03844506	4	4109441		Yes	-7.72
cg16558177	4	4109446		Yes	-2.25
cg11299854	5	132083184	CCNI2,KIF3A,SEPT8	Yes	43.6
cg05708550	5	137688227	CDC25C,FAM53C,KDM3B	Yes	8.4
cg16867657	6	11044877	ELOVL2	Yes	23.1
cg22736354	6	18122719	NHLRC1,TPMT	Yes	-15
cg14848772	6	27099813	HIST1H2AG,HIST1H2AH, HIST1H2BJ,HIST1H2BK,HIST1H4I	No	-14.3
cg15623062	6	31747133	Y_RNA.307	No	32.2
cg16489193	6	33240059		Yes	25.5
cg18468088	6	35490818	TULP1	No	5.96
cg04911280	6	44281184	AARS2,TCTE1,TMEM151B	Yes	-9.67
cg19291355	6	44281188	AARS2,TCTE1,TMEM151B	Yes	-5.16
cg05917988	6	44281197	AARS2,TCTE1,TMEM151B	Yes	2.77
cg20160885	7	5013524	MMD2,RBAK,RNF216L	Yes	49.6
cg19230755	7	65878503	GS1-124K5.2,GS1-124K5.6,	Yes	-2.26

[0173]

标记	染色体	位置	基因	CpG 岛	系数
			NCRNA00174,SKP1P1,TPST1, U6.862		
cg09941452	7	97557827		Yes	-14.2
cg26830108	7	100813299	AP1S1,C7orf52,VGF	No	1.22
cg19273773	7	102790112	NAPEPLD	Yes	1.56
cg14361627	7	130419116	KLF14	Yes	1.17
cg08097417	7	130419133	KLF14	Yes	53.5
cg02821342	7	130793551	MKLN1	No	-19.8
cg07392449	8	11324666	FAM167A	Yes	73
cg08318076	8	62051812	CLVS1	Yes	13
cg02560186	11	3602584	OR7E117P	Yes	27.8
cg08715791	11	66189297	MRPL11,NPAS4,SNORA43.2	Yes	-30.6
cg23156348	11	124981869	TMEM218	No	-21.1
cg10820926	14	30397408	PRKD1	Yes	19.2
cg06121469	15	44956098	PATL2,SPG11	No	-0.156
cg07477282	15	44956107	PATL2,SPG11	No	2.22
cg21801378	15	72612125	CELF6	Yes	-49.5
cg02331561	16	2391081	ABCA17P,ABCA3	Yes	-17
cg06144905	17	27369780	PIPOX	No	9.13
cg14692377	17	28562685	BLMH,SLC6A4,SNORD63.3	Yes	13.7
cg18569335	17	40171970	DNAJC7,NKIRAS2,ZNF385C	Yes	-30.6
cg26147554	18	712733	ENOSF1,YES1	Yes	17.7
cg21927946	19	4769688	C19orf30,MIR7-3	No	73.4
cg15789607	19	4769690	C19orf30,MIR7-3	No	-15.4
cg12589298	19	50828905	KCNC3,MYH14,NAPSB	Yes	15.3
cg06458239	19	58038573	MIR1274B,ZNF549,ZNF550, ZNF773	Yes	22.4
cg10729426	19	58038585	MIR1274B,ZNF549,ZNF550, ZNF773	Yes	13.5
cg26734668	19	58111094	ZIK1,ZNF134,ZNF530	No	9.2
cg22888484	20	37075185	SNHG11,SNORA39,SNORA60, SNORA71.3,SNORA71A, SNORA71C,SNORA71D	No	201

[0174] TCGA数据集还包括代表共319个肿瘤和匹配的正常组织样品(胸腺、肾、肺和皮肤)的甲基化图谱。有趣的是,我们的衰老模型的使用表明,在相同的个体中,肿瘤看上去比匹配的正常组织衰老多40%(威尔科克斯(Wilcoxon)测试, $p < 10^{-41}$, 图4D和4C)。不管主要组织类型如何,加速肿瘤衰老是明显的。我们通过检查所有70,387个与年龄相关的标记,研究了这是否是全球甲基化水平广泛变化的结果,其中,随着年龄的增加,44%倾向于增加而56%倾向于减少。在匹配的肿瘤和正常样品中的甲基化分值支持了一个发现,不管趋势方向如何,肿瘤与74%标记的年长值相吻合(二项式 $p \sim 0$)。而且,构建于匹配的正常和肿瘤样品的分离衰老模型确认了显著地衰老效应(表S4)。

[0175] 不同的衰老速率导致不同的甲基化

[0176] 如果个体年龄确实以不同速率,可以预料他们个体甲基化随着时间的推移应该不同。这基于前提是较年轻的甲基化共享某些相似性,及随着时间的推移个体积累的变化,这些相似性减少。在同卵双生中已观察到这种所谓表观遗传漂移的效应(Fraga et al., 2005),但是,很少提出特定的假说对此作出解释。为了检查我们样品中的表观遗传漂移,我们从预期总体平均数计算了每个甲基化标记的偏移值作为其平方距离(图5A)。然后,除了测试其甲基化分值随着年龄的变化而变化的标记(图5B和5C),我们能测试其变异随着年龄的变化而变化的标记(图5D和5C)(Breusch and Pagan,1979)。渐增的变异是一种普遍的现象-我们鉴别了27800个标记,标记的变异与年龄有显著的相关性($FDR < 0.05$),其中27,737 (99.8%)代表增加而不是减少的变异(图5E和5F)。对于任何给定的个体,尤其高或低的甲基化变异是衰老速率的强预测器($R = 0.47, p \sim 0$),表明了衰老速率的差别为部分甲基化异质和表观遗传漂移作出了解释。

[0177] 检查表观遗传漂移的另一种方法是依据Shannon熵,或者是随着时间的推移甲基化信息内容的缺失(Shannon and Weaver,1963)。CpG标记熵的增加意味着其甲基化状态在细胞群中变得难以预测,如其甲基化分值趋向于50%。确实,与样品行列中甲基化分值的变化相关的所有标记,70%趋向于甲基化分值为50%(图6A,二项式 $p \sim 0$,表S2)。所以,我们观察到在样品行列中甲基化熵的高显著的增加($R = 0.21, p < 10^{-7}$)。而且,基于AMAR,个体极端甲基化熵与加速的衰老速率具有很高的关联性($R = 0.49, p \sim 0$,图6B)。

[0178] 衰老速率和转录组

[0179] 由于甲基化中的变化与基因表达的变化有直接的关联(Sun et al.,2011),我们感兴趣的是,在衰老甲基化中的这些变化是否反映在人类转录组中功能层面上及是否反映了衰老速率的差别。为此,我们从488个个体,跨越年龄范围为20至75的全血中获得和分析公开的可用的基因表达图谱(Emilsson et al.,2008)。我们找到了与年龄相关的基因表达(326个基因, $FDR < 0.05$)和渐增的表达变异的基因的有利的证据(二项式 $p < 10^{-276}$)。引人瞩目的是,我们发现,在我们的数据中,与年龄相关的基因表达图谱更可能具有附近的与年龄相关的甲基化标记($p < 0.01$,表S6)。基于在甲基化中与年龄相关的基因表达,我们利用这些信息构建了一个衰老模型(图7A,表S7)。这个模型显示了使用表达数据测量衰老速率明显的的能力,与女人相比,重现了我们在与女人相比时男人渐增的衰老速率的发现(图7B,11%差异, $p < 10^{-4}$)。性别效应在所构建的模型中是不存在,使用所有的可利用的基因而不是那些在甲基化中与年龄相关的变化($p > 0.05$)。因此,与年龄相关的变化表明基因表达模式的功能变化。

[0180] 表S6,与图7有关的在甲基化和转录组两者中的与衰老相关的基因

[0181] 映射到年龄相关的甲基化标记和表转录组中与年龄相关的变化的基因列表。

[0182]

ABCA3	BFSP1	CDKN1C	DEPDC7	FCGBP	IL7R	MAN1C1
ABCB9	BHLHE40	CEBPG	DGKA	FGFBP2	INPP4B	MB21D2
ABLIM1	BLNK	CECR5	DLL1	FGFRL1	IRS1	MEOX1
ACAA2	BYSL	CENPE	DNASE1L3	FLNB	ITFG2	MEST
ACCN2	C10orf128	CENPV	DNMT3A	FOXP1	ITGA6	MLF1
ACSF2	C12orf23	CHMP7	DNMT3B	FZD1	ITM2C	MPI
ACVR2A	C16orf45	CHSY3	DPH5	GAL3ST4	JAKMIP1	MS4A3
AEBP1	C17orf58	CIAPIN1	DUSP4	GATA3	KAT2A	MS4A4A
AGBL2	C1orf172	CISH	DYNLL1	GF11	KATNAL1	MT1E
AGPAT4	C1orf21	CMC1	ECT2	GLT25D2	KCNMB4	MT1M
AK5	C1orf216	COBLL1	EDAR	GNG7	KIAA1841	MTSS1
ALDH5A1	C1orf51	COL5A3	EEF1G	GPC2	KLF4	MTUS1

[0183]

ANKRD13B	C21orf63	CR2	EFNA1	GPR114	KLF6	MXRA8
ANKS6	C2orf40	CRIP1	EOMES	GPR137B	KLHL14	MYC
ANXA1	C6orf97	CRTAM	EPHA1	GPR153	KLHL3	MYO6
APBA2	CACHD1	CRTC3	EPHA2	GPR56	KLRG1	MYOF
APBB1	CACNA2D2	CSF1R	EPHX2	GSC	LAMA5	NBEA
APOBEC3H	CALHM2	CST7	EPPK1	GTSF1	LBH	NCAPH
ARAP2	CAMK2N1	CTLA4	EXPH5	GYGI	LDLRAP1	NEFH
ARHGEF4	CAPN2	CTNNA1	FAIM3	GZMH	LEF1	NELL2
ATPIB1	CCDC106	CTSL1	FAM129C	HEXIM1	LGALS1	NHLRC1
B3GAT1	CCR10	CX3CR1	FAM134B	HIST1H3D	LILRA4	NKG7
BACH2	CCR7	CYP2J2	FAM13A	HOPX	LIMS2	NMT2
BATF3	CD200	CYP4F12	FANI	IGFBP7	LMO7	NMUR1
BCAS4	CD244	DCBLD2	FASLG	IGLL1	LPCAT1	NOB1
BCL7A	CD8B	DDB2	FBL	IL10RA	LRP11	NOP16
BCL9	CD9	DEFA4	FBLN2	IL16	LRRC32	NOSIP
BDHI	CDCA7L	DENND2D	FBXO24	IL4I1	LTK	NPM3
NRCAM	PDGFRB	PRSS23	RHOC	SLC27A5	TCF7	VCAM1
NSUN5	PEL13	PRSS35	RNASE2	SLC2A6	TCF7L2	VIT
NT5E	PHGDH	PTGDS	RNASE3	SLC45A3	TGFR3	WARS
NTAN1	PHLDA3	PTGER2	RNF144A	SOCS2	TIGIT	ZBED3
NUAK1	PHYHD1	PTPRK	ROBO1	SORCS3	TM6SF1	ZFYVE28
OSBPL10	PI16	PTTG1	RPL13	SOX15	TMEM121	ZNF135
OXNAD1	PIK3IP1	PUS1	RUNX3	SPEG	TMEM8B	ZNF167
P2RX5	PLAG1	PYROXD1	S100A10	SPIB	TMIGD2	ZNF177
PACSLN1	PLCG1	RAB15	S1PR5	SPINK2	TNFRSF17	ZNF263
PALLD	PLEKHA7	RAB27B	SATB1	SPN	TNFRSF25	ZNF285
PAQR4	PLXDC1	RAB6B	SCARB1	STOM	TPPP3	ZNF365
PCBP4	PMEPA1	RAB6C	SEC14L2	STX8	TRAF5	ZNF462
PCDH12	PMP22	RAD54B	SEMA3G	SUSD1	TRAP1	ZNF528
PCSK4	POMC	RAMP1	SFRP5	SYT11	TRIM2	ZNF544
PCSK5	POU2AF1	RAPGEF6	SFTPD	TARBP1	TSPAN13	ZNF551
PDE6B	PPAP2C	RASD1	SIRPG	TBX21	TSPAN2	ZSCAN18
PDE7A	PPM1J	RASGEF1A	SLAMF7	TCAP	TWIST1	
PDE9A	PPP2R2B	RGMA	SLC1A7	TCF3	TXNDC5	
PDGFD	PRR5L	RGS9	SLC23A1	TCF4	USP18	

[0184] 表S7,与图7有关的转录衰老模型

[0185] 基于转录数据用于预测年龄的基因列表和系数。

[0186]

基因	系数
ABLIM1	-4.537363687
ACCN2	-4.021935755
ACVR2A	5.862922173
AK5	-10.29726151
ANXA1	6.307730249
ASNS	23.66865779
AUTS2	-13.3985662
C16orf45	4.553248948
CACHD1	-7.768187189
CDKN1C	-0.0105012
CENPV	-2.462314825
CMC1	8.866490009
CR2	-1.78645877
CRIP1	4.33558575
EFNA1	-5.741766145
EPHA2	2.917895917
FAIM3	1.019538625
TNFRSF17	-6.154501401

基因	系数
FBLN2	-2.061520114
FLNB	-3.844863485
FZD1	1.375051746
GPC2	-7.431385678
GSC	4.904090057
GTSF1	12.58953522
HIST1H3D	-8.692565907
IGLL1	-4.235566899
KRT72	2.814127932
LRP11	5.664133584
MEOX1	13.8516364
MS4A3	-2.661573104
NEFH	-6.728594491
NMT2	-15.38338708
NOSIP	-13.61680769
NT5E	-1.994658678
PHLDA3	18.71229769
TXNDC5	-0.265342977

基因	系数
PHYHD1	-5.052538719
PLXDC1	-5.337661458
POMC	-6.100365433
PRSS35	-6.498528559
PTGER2	-8.407414661
PYROXD1	13.70056157
RGMA	5.458322024
ROBO1	-7.342718162
SEC14L2	-2.887682148
SFRP5	4.430923586
SLC45A3	8.799140451
SORCS3	9.998064269
SPEG	0.574659287
SPINK2	0.302316458
SYT11	7.093819787
TMEM8B	-13.0069907
TMIGD2	6.006761191
ZNF285	-4.729710661

[0187] 结论

[0188] 在本研究中,我们已经表明全基因组甲基化模式代表生物衰老速率的强大的并可复制的生物标记。这些模式能使衰老甲基化的定量模型显示高精度度及区分衰老中包括性别和基因变异的相关的因素的能力。而且,我们把这种模型应用到多种组织的能力表明了一个常见的分子钟的能力,通过甲基化变化的部分调节。这些变化是否在跨细胞群中细胞内发生的水平一致或者经过一段时间之后,是否在组织成分中随时间推移显示一致的变化还有待观察。

[0189] 从全血中预测年龄的能力可在纵向研究中允许广泛的分析,如弗雷明汉(Framingham)研究、妇女健康倡议、从新生儿格思里(Guthrie)卡收集的的血样品和具有丰富注释的生物统计学和疾病特征的其他纵向研究。衰老趋势可从具有许多潜在的现实意义的此类研究中显现出来,从健康评估和疾病预防到法医分析。类似于本研究中的性别效应,其他影响AMAR的生物或如吸烟、饮酒或节食的环境因素的鉴别可允许对健康和长寿影响的定量评估。一个有用的例子可定期地评估使用AMAR的个体衰老速率及确定节食或环境因素是否可加速或延缓衰老过程和如年龄相关性黄斑变性的疾病。作为人类衰老模型的改进,可以想象的是,有一天,用分子谱测量的生物年龄在临床评估和患者治疗中可取代实足年龄。

[0190] 实施例1的参考文献

[0191] Alisch, R.S., Barwick, B.G., Chopra, P., Myrick, L.K., Satten, G.A., Conneely, K.N., and Warren, S.T. (2012). Age-associated DNA methylation in pediatric populations. *Genome Res.* 22, 623-632.

[0192] Atzmon, G., Rincon, M., Schechter, C.B., Shuldiner, A.R., Lipton, R.B., Bergman, A., and Barzilai, N. (2006). Lipoprotein genotype and conserved pathway for exceptional longevity in humans. *PLoS Biol.* 4, e113.

- [0193] Austad, S.N. (2006). Why women live longer than men: sex differences in longevity. *Gend. Med.* 3, 79–92.
- [0194] Barres, R., and Zierath, J.R. (2011). DNA methylation in metabolic disorders. *Am. J. Clin. Nutr.* 93, 897S–900.
- [0195] Bell, J.T., Pai, A.A., Pickrell, J.K., Gaffney, D.J., Pique-Regi, R., Degner, J.F., Gilad, Y., and Pritchard, J.K. (2011). DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines. *Genome Biol.* 12, R10. <http://www.ncbi.nlm.nih.gov/pubmed/21251332>.
- [0196] Bell, J.T., Tsai, P.-C., Yang, T.-P., Pidsley, R., Nisbet, J., Glass, D., Mangino, M., Zhai, G., Zhang, F., Valdes, A., et al.; MuTHER Consortium. (2012). Epigenome-wide scans identify differentially methylated regions for age and age-related phenotypes in a healthy ageing population. *PLoS Genet.* 8, e1002629.
- [0197] Bellacosa, A., Cicchilitti, L., Schepis, F., Riccio, A., Yeung, A.T., Matsumoto, Y., Gojemis, E.A., Genuardi, M., and Neri, G. (1999). MED1, a novel human methyl-CpG-binding endonuclease, interacts with DNA mismatch repair protein MLH1. *Proc. Natl. Acad. Sci. USA* 96, 3969–3974.
- [0198] Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. B* 57, 289–300. Bertoni, C., Rustagi, A., and Rando, T.A. (2009). Enhanced gene repair mediated by methyl-CpG-modified single-stranded oligonucleotides. *Nucleic Acids Res.* 37, 7468–7482.
- [0199] Bibikova, M., Barnes, B., Tsan, C., Ho, V., Klitzke, B., Le, J.M., DeIano, D., Zhang, L., Schroth, G.P., Gunderson, K.L., et al. (2011). High density DNA methylation array with single CpG site resolution. *Genomics* 98, 288–295.
- [0200] Blair, S.N., Koh, H.W., 3rd, Paffenbarger, R.S.J., Jr., Clark, D.G., Cooper, K.H., and Gibbons, L.W. (1989). Physical fitness and all-cause mortality. A prospective study of healthy men and women. *JAMA* 262, 2395–2401.
- [0201] Bocklandt, S., Lin, W., Sehl, M.E., Sánchez, F.J., Sinsheimer, J.S., Horvath, S., and Vilain, E. (2011). Epigenetic predictor of age. *PLoS ONE* 6, e14821.
- [0202] Boks, M.P., Derks, E.M., Weisenberger, D.J., Strengman, E., Janson, E., Sommer, I.E., Kahn, R.S., and Ophoff, R.A. (2009). The relationship of DNA methylation with age, gender and genotype in twins and healthy controls. *PLoS ONE* 4, e6767.
- [0203] Bohlatti, V., Schwartz, J., Wright, R., Litonjua, A., Tarantini, L., Suh, H., Sparrow, D., Vokonas, P., and Baccarelli, A. (2009). Decline in genomic DNA methylation through aging in a cohort of elderly subjects. *Mech. Ageing Dev.* 130, 234–239.
- [0204] Breusch, T.S., and Pagan, A.R. (1979). A Simple Test for Heteroscedasticity and Random Coefficient Variation. *Econometrica* 47, 1287.
- [0205] Christensen, B.C., Houseman, E.A., Marsit, C.J., Zheng, S., Wrensch, M.R.,

- WiemeIs, J.L., Nelson, H.H., Karagas, M.R., Padbury, J.F., Bueno, R., et al. (2009). Aging and environmental exposures alter tissue-specific DNA methylation dependent upon CpG island context. *PLoS Genet.* 5, e1000602.
- [0206] Collins, F.S., and Barker, A.D. (2007). Mapping the cancer genome. Pinpointing the genes involved in cancer will help chart a new course across the complex landscape of human malignancies. *Sci. Am.* 296, 50–57.
- [0207] deMagalhães, 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.
- [0208] Emilsson, V., Thorleifsson, G., Zhang, B., Leonardson, A.S., Zink, F., Zhu, J., Carlson, S., Helgason, A., Walters, G.B., Gunnarsdottir, S., et al. (2008). Genetics of gene expression and its effect on disease. *Nature* 452, 423–428.
- [0209] EpeI, E.S., Blackburn, E.H., Lin, J., Dhabhar, F.S., Adler, N.E., Morrow, J.D., and Cawthon, R.M. (2004). Accelerated telomere shortening in response to life stress. *Proc. Natl. Acad. Sci. USA* 101, 17312–17315.
- [0210] Esteller, M. (2008). Epigenetics in cancer. *N. Engl. J. Med.* 358, 1148–1159.
- [0211] Feinberg, A.P., Irizarry, R.A., Fradin, D., Aryee, M.J., Murakami, P., Aspelund, T., Eiriksdottir, G., Harris, T.B., Launer, L., Gudnason, V., and Fallin, M.D. (2010). Personalized epigenomic signatures that are stable over time and covary with body mass index. *Sci. Transl. Med.* 2, 49ra67.
- [0212] Fraga, M.F., and Esteller, M. (2007). Epigenetics and aging: the targets and the marks. *Trends Genet.* 23, 413–418.
- [0213] Fraga, M.F., Balléstar, E., Paz, M.F., Ropero, S., Setien, F., Balléstar, M.L., Heine-Suñer, D., Cigudosa, J.C., Urioste, M., Benitez, J., et al. (2005). Epigenetic differences arise during the lifetime of monozygotic twins. *Proc. Natl. Acad. Sci. USA* 102, 10604–10609.
- [0214] Fraser, H.B., Khaitovich, P., Plotkin, J.B., Pääbo, S., and Eisen, M.B. (2005). Aging and gene expression in the primate brain. *PLoS Biol.* 3, e274.
- [0215] Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. *J. Stat. Softw.* 33, 1–22.
- [0216] Gentleman, R.C., Carey, V.J., Bates, D.M., BoIstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., et al. (2004). Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* 5, R80.
- [0217] Harley, C.B., Futcher, A.B., and Greider, C.W. (1990). Telomeres shorten during ageing of human fibroblasts. *Nature* 345, 458–460.
- [0218] Hartwig, A., and SchIepgreII, R. (1995). Induction of oxidative DNA damage by ferric iron in mammalian cells. *Carcinogenesis* 16, 3009–3013.

- [0219] Heyn, H., Li, N., Ferreira, H. J., Moran, S., Pisano, D. G., Gomez, A., Diez, J., Sanchez-Mut, J. V., Setien, F., Carmona, F. J., et al. (2012). Distinct DNA methylomes of newborns and centenarians. *Proc. Natl. Acad. Sci. USA* 109, 10522–10527.
- [0220] Jones, P. A., and Laird, P. W. (1999). Cancer epigenetics comes of age. *Nat. Genet.* 21, 163–167.
- [0221] Karthikeyan, G., Lewis, L. K., and Resnick, M. A. (2002). The mitochondrial protein frataxin prevents nuclear damage. *Hum. Mol. Genet.* 11, 1351–1362.
- [0222] Kim, E. J., Kim, M.-K., Jin, X.-J., Oh, J.-H., Kim, J. E., and Chung, J. H. (2010). Skin aging and photoaging alter fatty acids composition, including 11, 14, 17-eicosatrienoic acid, in the epidermis of human skin. *J. Korean Med. Sci.* 25, 980–983.
- [0223] Lane, M. A., Baer, D. J., Rumpier, W. V., Weindruch, R., Ingram, D. K., Tilmon, E. M., Cutler, R. G., and Roth, G. S. (1996). Calorie restriction lowers body temperature in rhesus monkeys, consistent with a postulated anti-aging mechanism in rodents. *Proc. Natl. Acad. Sci. USA* 93, 4159–4164.
- [0224] Li, R., Li, Y., Kristiansen, K., and Wang, J. (2008). SOAP: short oligonucleotide alignment program. *Bioinformatics* 24, 713–714.
- [0225] Moniz, L., Dutt, P., Haider, N., and Stambolic, V. (2011). Nek family of kinases in cell cycle, checkpoint control and cancer. *Cell Div.* 6, 18.
- [0226] Murgatroyd, C., Patchev, A. V., Wu, Y., Micalé, V., Bockmühl, Y., Fischer, D., HoIsboer, F., Wotjak, C. T., Almeida, O. F. X., and Spengler, D. (2009). Dynamic DNA methylation programs persistent adverse effects of early-life stress. *Nat. Neurosci.* 12, 1559–1566.
- [0227] Ohgami, R. S., Campagna, D. R., McDonald, A., and Fleming, M. D. (2006). The Steap proteins are metallo-reductases. *Blood* 108, 1388–1394.
- [0228] Rakyan, V. K., Down, T. A., Maslau, S., Andrew, T., Yang, T.-P., Beyan, H., Whittaker, P., McCann, O. T., Finer, S., Valdes, A. M., et al. (2010). Human aging-associated DNA hypermethylation occurs preferentially at bivalent chromatin domains. *Genome Res.* 20, 434–439.
- [0229] Saito, T., Iwata, N., Tsubuki, S., Takaki, Y., Takano, J., Huang, S.-M., Suemoto, T., Higuchi, M., and Saido, T. C. (2005). Somatostatin regulates brain amyloid beta peptide Abeta42 through modulation of proteolytic degradation. *Nat. Med.* 11, 434–439.
- [0230] Shannon, C. E., and Weaver, W. (1963). *The Mathematical Theory of Communication* (Champaign, IL: University of Illinois Press).
- [0231] Small, K. S., Hedman, A. K., Grundberg, E., Nica, A. C., Thorleifsson, G., Kong, A., Thorsteindottir, U., Shin, S.-Y., Richards, H. B., Soranzo, N., et al.; GIANT Consortium; MAGIC Investigators; DIAGRAM Consortium; MuTHER Consortium. (2011). Identification of an imprinted master trans regulator at the KLF14 locus

related to multiple metabolic phenotypes. *Nat. Genet.* 43, 561–564.

[0232] Suh, Y., Atzmon, G., Cho, M.-O., Hwang, D., Liu, B., Leahy, D. J., Barzilai, N., and Cohen, P. (2008). Functionally significant insulin-like growth factor I receptor mutations in centenarians. *Proc. Natl. Acad. Sci. USA* 105, 3438–3442.

[0233] Sun, Z., Asmann, Y. W., Kalari, K. R., Bot, B., Eckel-Passow, J. E., Baker, T. R., Carr, J. M., Khrebtukova, I., Luo, S., Zhang, L., et al. (2011). Integrated analysis of gene expression, CpG island methylation, and gene copy number in breast cancer cells by deep sequencing. *PLoS ONE* 6, e17490.

[0234] Tatar, M., Bartke, A., and Antebi, A. (2003). The endocrine regulation of aging by insulin-like signals. *Science* 299, 1346–1351.

[0235] Troyanskaya, O., Cantor, M., Sherlock, G., Brown, P., Hastie, T., Tibshirani, R., Botstein, D., and Altman, R. B. (2001). Missing value estimation methods for DNA microarrays. *Bioinformatics* 17, 520–525.

[0236] Valdes, A. M., Andrew, T., Gardner, J. P., Kimura, M., Oelsner, E., Cherkas, L. F., Aviv, A., and Spector, T. D. (2005). Obesity, cigarette smoking, and telomere length in women. *Lancet* 366, 662–664.

[0237] Vijg, J., and Campisi, J. (2008). Puzzles, promises and a cure for ageing. *Nature* 454, 1065–1071.

[0238] Wheeler, H. E., Metter, E. J., Tanaka, T., Absher, D., Higgins, J., Zahn, J. M., Wilhelm, J., Davis, R. W., Singleton, A., Myers, R. M., et al. (2009). Sequential use of transcriptional profiling, expression quantitative trait mapping, and gene association implicates MMP20 in human kidney aging. *PLoS Genet.* 5, e1000685.

[0239] Willcox, B. J., Donlon, T. A., He, Q., Chen, R., Grove, J. S., Yano, K., Masaki, K. H., Willcox, D. C., Rodriguez, B., and Curb, J. D. (2008). FOXO3A genotype is strongly associated with human longevity. *Proc. Natl. Acad. Sci. USA* 105, 13987–13992.

[0240] Xiong, M., Dong, H., Siu, H., Peng, G., Wang, Y., and Jin, L. (2010). Genome-Wide Association Studies of Copy Number Variation in Glioblastoma. *Proceedings of the 4th International Conference on Bioinformatics and Biomedical Engineering (iCBBE)*, 1–4.

[0241] Yacubova, E., and Komuro, H. (2002). Stage-specific control of neuronal migration by somatostatin. *Nature* 415, 77–81.

[0242] 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.

[0243] Zou, H., and Hastie, T. (2005). Regularization and variable selection via the elastic net.

[0244] *J. R. Stat. Soc. Series B Stat. Methodol.* 67, 301–320.

[0245] 实施例2

[0246] 构建衰老甲基化模型

[0247] 我们测量在从302个高加索人全血样品收集的基因组DNA中的485,577 CpG标记的甲基化状态。其中,80个个体被诊断有2-型糖尿病,其中的22个也具有糖尿病性肾病。为了进一步的研究和验证,获得了第二行列的甲基化状态,包括188个西班牙裔人。第二群体的每个人被诊断有2-型糖尿病及89个人也具有糖尿病性肾病。进行仔细筛选和正常化以除去性别、批次或其他未知的协变量的影响。

[0248] 一般来说,我们假设,生物活性跟踪实足年龄,允许我们从实足年龄中推断生物模型。我们假定,分子活性中从共同基线的变化将反映衰老过程的减缓或加速,其中,疾病、环境和遗传学可能作出贡献。

[0249] 形式上,我们把生物年龄(bioage)定义为:

$$[0250] \quad Bioage = f(M_{sub}) = Age + \sum_{j=1}^f \alpha_j c_j + \epsilon$$

[0251] 其中, M_{sub} 是甲基化数据小子集, α_j 是数值系数, c_j 是第j个性状,及 ϵ 是模型误差。这儿的临界点为通过选择与实足衰老协调关联的模型探针,相对于这些相同的探针甲基状态的改变可能反映为不是衰老过程的衰减就是衰老过程放大。

[0252] 我们考虑到,生物衰老的速率不是一成不变的,及在人类发展和衰老的不同里程碑中,在生物衰老中将发生大的变化。通过首先使用单变量关联测试,我们测试了这一假说,以鉴别主要行列中顶端年龄相关的甲基化标记($FDR < 0.05$)。然后,我们使用熵尺度测定年轻和年老个体之间的那些标记的相对一致性(图1)。我们发现,相关标记在年轻个体中比在年老个体中更一致($p < 0.05$)。这表明甲基化衰老模式与年轻个体相似,但随着时间偏离。

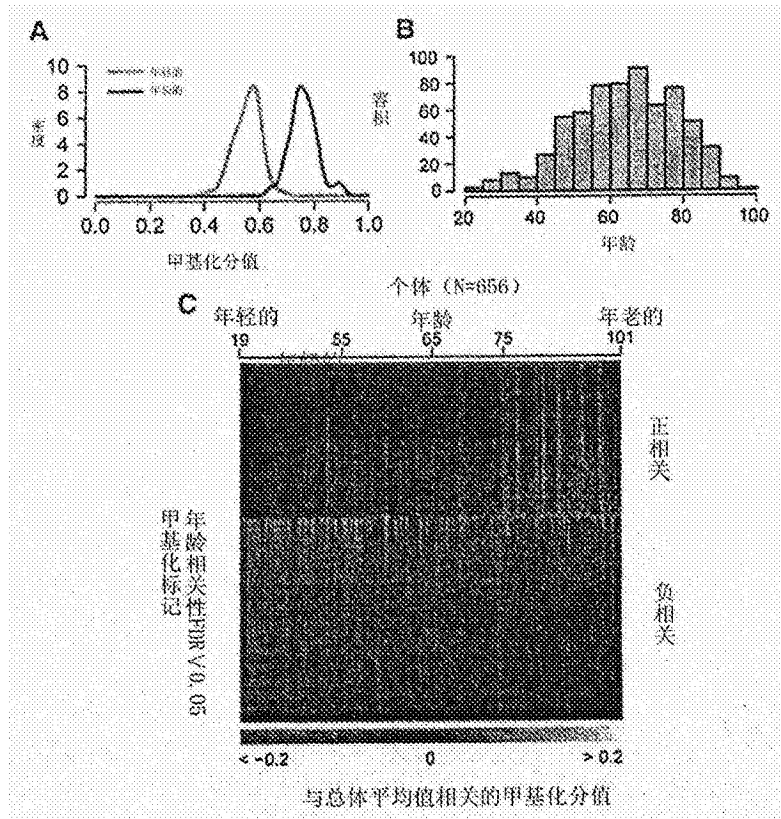


图1

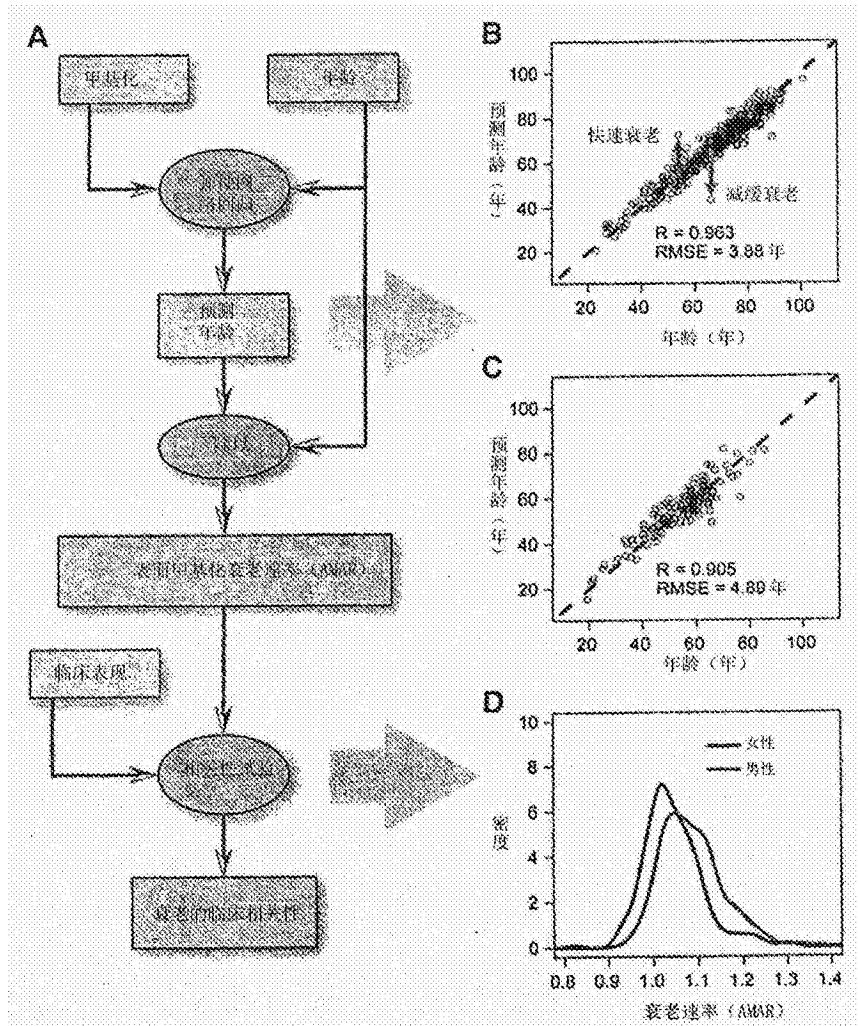


图2

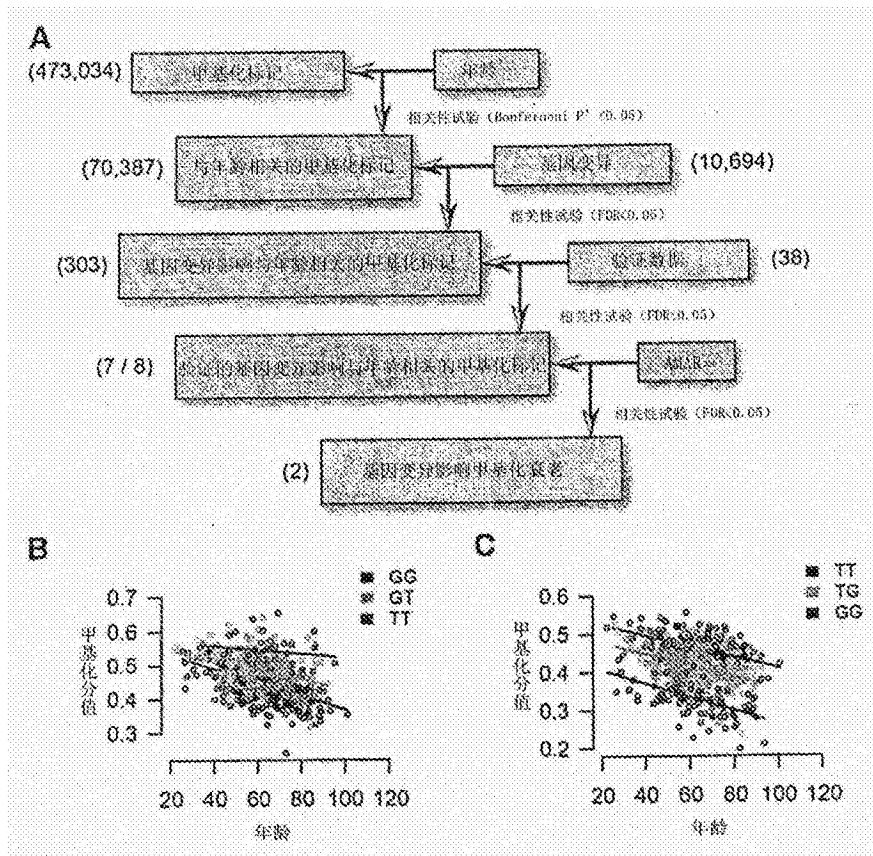


图3

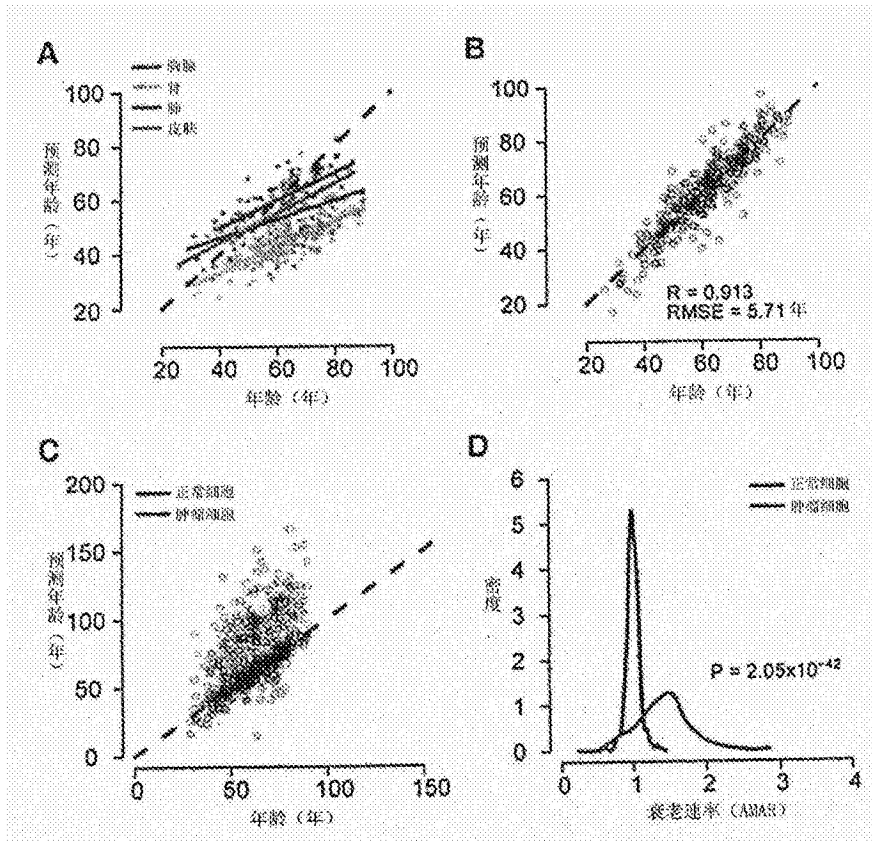


图4

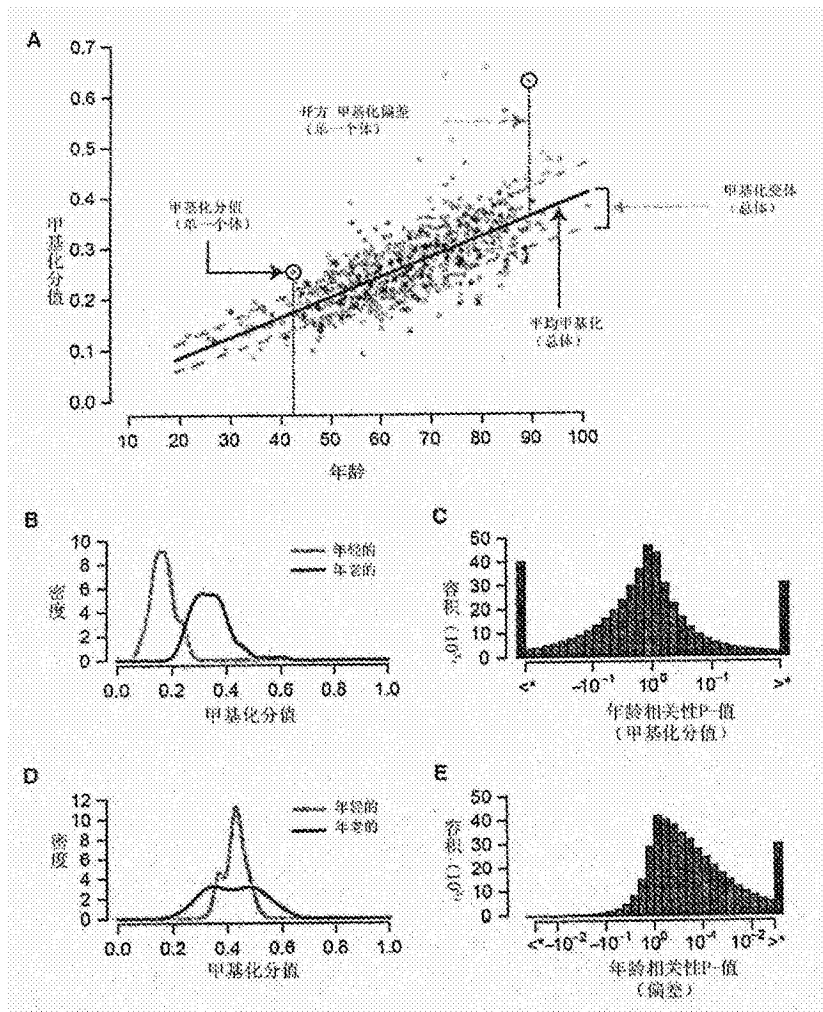


图5

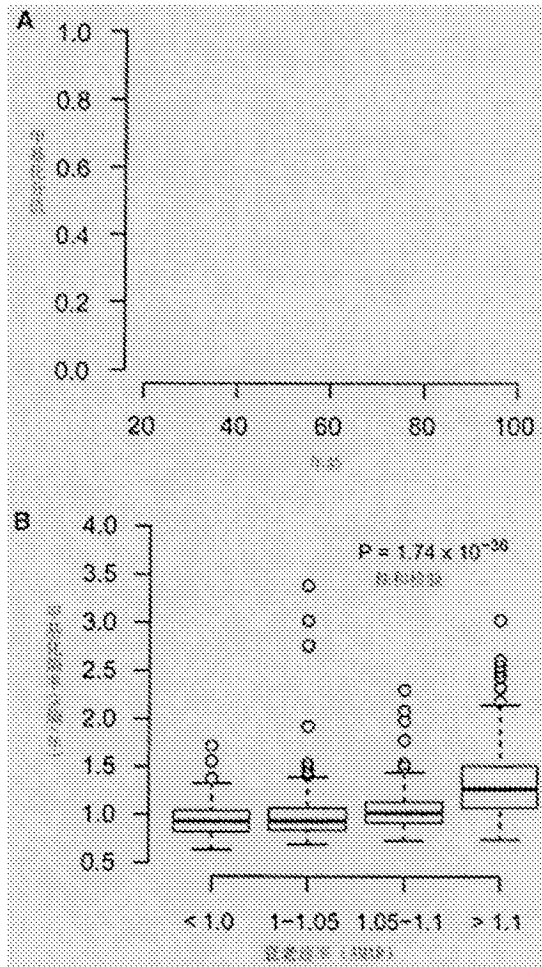


图6

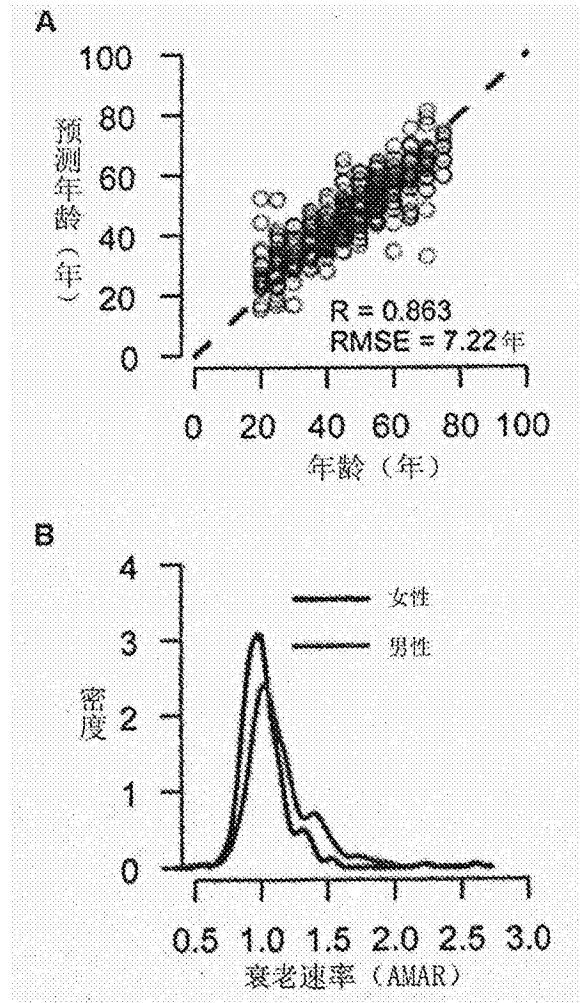


图7

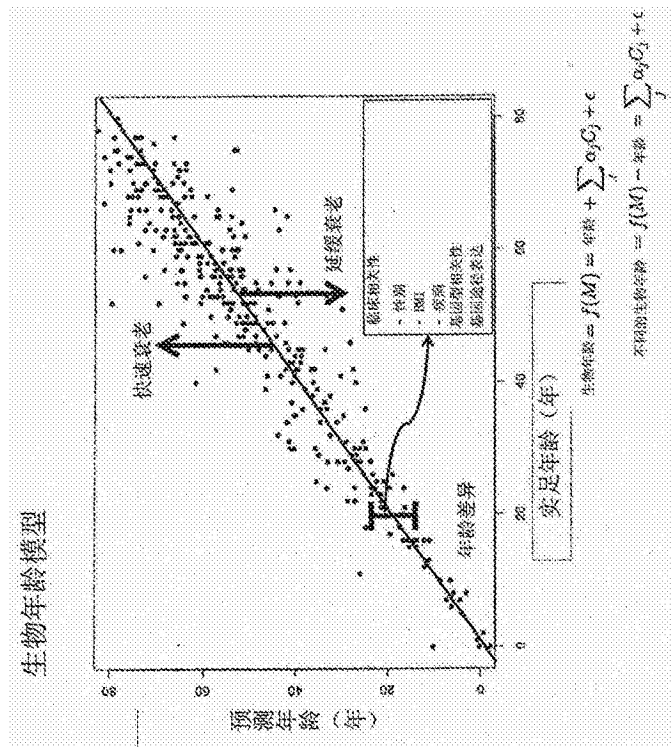


图8

ID	名称	地址 A_ID	序列基序 A_序列序列	地址 B_ID
19934	cg04400972	74745443	AAAAATAAAACATCTCAAACCTCAGATTTACAAAAAC	13601340
28973	cg09809672	12737501	CAATTCAAAAACCA TTTCATCTAAAAAATTTAACTCTAACCAAAACCA	NA
30213	cg10501210	43774458	ACRAACATCTCTC CAAACTACAACACCTTACAACAAAAACAAAATTT	66775341
39647	cg16054275	17771388	AAATCTACAACA TCCAAATATAAATTAACCCCAAAACCACTCAA	NA
47340	cg20822990	50699474	TCAATAATAACRAAC CAGAAAAAATCTACRAAAATCAGACAAAAAAAT	NA
49956	cg22512670	70754418	ACTTATCAACTAAAC ATAATAACAACACTACTAATACCCATCTTAACCTAAAA	NA
54990	cg25410668	26779422	CACTAAAATCRCC ACACGCRATAAAATTAATTTACTTTTTTACATAAAAA	NA
61752	cg02085953	66719449	ATCAGATCTCTTC CTAAAATTTCAAACCTAAAAACCCCTAAAAATAAACTCC	NA
67888	cg06639320	13732383	TCACCTAAAAAAC CAAAACTCTTTCTTCTRTACCTCCRAATCTTAAAA	NA
86462	cg22016779	10606390	ACACAATAATTATC RCAAAACCCCTTAAAAAACCACCAATACAACAAT	NA
86644	cg22158769	17654495	AATTAACCAAAACAC ACAAAAAACTTTTAAAAAACAACCTCAACAAC	68643399
86991	cg22454769	66727363	TAACTTACCATA GTTAAAAACACAATAAATTAATCAAAAAACATCACCTC	23744359
88430	cg23606718	20628466	CAACATAAATCTCA TCTCAAAACCTTAAACCACTTACCCTAAAAACCCAC	58618322
88984	cg24079702	47601467	AATACAACAACA AAAAACAATAATTAACAATAATCAACCTCAAC	50810372
93051	ch.2.30415474F	51667506	ATAAACTCTCAACA AAAACTAACTCTTCTTACCACATATAATCATCAA	NA
	ch.2.30415474F		CAAACTCTATAAT	

图 9A

图 9A

ID	名称	地址_A_ID	序列图 A_ 探针序列	地址_B_ID
93891	cg00481951	14709443	CRITTCACACCTTAAATCAACRCTTCCAAAATCAA CACAAAAATAAAC	NA
97218	cg03607117	70676435	TATCTACAACCCRAAATAAACAACRCAAACTAAATCR AAACTAAACCAAAAC	NA
98073	cg04474832	31636354	TAAACRITTTACTCTTACTTACTCTACAAAAATTA ACAACATCTAACCC	NA
100937	cg07553761	43625429	ATATAACACGCAAAATAAAAAATAATTCTCAAAAAAC CATCAACCCCAACA	17688430
116505	cg25478614	60721480	AAAAACCAAAAAACCAAAAACTCTTTAAAAAAA CTAAACATCCCTTAC	NA
120672	cg02650266	66683464	ACTATCTCAAAAAACCAAAAAATACTAAAAAAA CAACAACAACAACA	73675449
137234	cg25428494	68769372	AAATATATTACATTTATAATAACATCTAAACTTTAA TATACTATCACCC	NA
146444	cg08234504	24749379	TAAAACTAATTCACCTATTACACACAAAAACAT AACTACCAATAAAC	NA
159652	cg23500537	48779348	TACAACACAGATCCAAAACCTAACAAAAACRAACAC TCTACCAAACTCAC	NA
164052	cg00486113	37770376	AAACCTTAAACCCCAACCCCAAAAAATAACAAATAT AAAAATCCCTCAACC	NA
172686	cg06493994	49740386	CAAAAAATACRATAAAAAATACTTCCAAAAATTA TCTAAATCTTCRC	NA
172936	cg06685111	59773369	TTACCAATCTAAATCCRTCTTAATACCTTACTATAC ATACAATCTACTC	NA
181162	cg13001142	74697332	CACCRITCAAAAAATTAACAAATTCAAATCTAACCT AATCAAAATTTATC	NA
186168	cg16867657	57761335	TCTAAACAACAATAAATAATTTCTAAAACTCCATA AACATTAACCCACA	17685408
190156	cg20052760	32659462	ATTTTCACRCCATCAACAAITTTATTTCTCACCTAA CTRAAAAAACAATC	NA

图 9B

图 9B

ID	名称	地址_A_ID	等位基因A_测序序列	地址_B_ID
193374	cg22736354	26771442	ACTCRAATACAAAATATACTTTAAAAAATTTAAACCA CCRACCAACACBRAC	NA
204211	cg03473532	31780418	AAITTCAAAATTTTCAAAATAATAATAAACAACACTACTA TCTCAACACATAC	NA
209326	cg07927379	42797400	ACTAAAACACAAAAACATACAAACACCCCAACCTC ATATAATAAAAAACCA	27785403
209350	cg07955595	49679509	ATCTCCACATTTCTTCTCTACCAACAAACCAAAATA ATAATAACAAAACA	57684340
209501	cg08097417	24706423	ATATTTAACAACTCAAAAATTTATCTTCTCRCR TTCTTTCTCTACC	NA
209919	cg08540945	24789307	AAITTCAAAACCTTTAAACCCCAACACTCTACAAC AAAAACAACAACA	59640388
215998	cg14361627	42786323	RACCCCCRACTAAATCATATTTAACAACTCAAAA ATTATCTTATCTCC	NA
222324	cg20426994	55807502	ATAACACTTAACACAAATATAACAACTCTCTCC AAAAACACTAAATCCA	12608480
224209	cg22285878	49781392	TCTTCTACCAACAACCAATAATAACAACAAC AAAACTCCCAATCA	49684508
236164	cg07583137	23646337	ACAACCCCA TTTAAAA TTTTAAATACACTAAAAATC ATTCAAAAATCTC	NA
242497	cg16419235	72778377	CAAAAACTAA TTTTACAATCACTAAAACCTACA ACAACACTCTCCA	26722332
279142	cg19935065	27739481	TATTTTAAAACTTTTACTATCTCAAAAACACRITA AACACTCACAACC	NA
280381	cg22796704	60625432	CRAAAAAACCRCA TTTCAAAAACACTTACTCTCTAA ATACCTAATAAATC	NA
286836	cg02046143	38797338	TTTCCACACATACTCTCTACTTATTACTCTAA ACCAAAAAAACA	15671301
290175	cg04940570	46668492	ACACACCTCAACACTTAAACAAAATAACACTAA AAACCAAAAATCCA	30738394

图 9C

图 9C

ID	名称	地址_A_ID	单位基因A_操作序列	地址_B_ID
291937	cg06419846	40649319	TAATCAGATATCTCCACTACACTAAAACTTCCA ACTAACAAACAACA	16728494
296834	cg11067179	41621317	CRAAAAATACTCCAGCCRCGTATATAAACACAAT AAATACCAAAITACC	NA
308206	cg22213242	25756359	ATCTACTAACACACACCAACAATCTAACACTCATCT ATATCCACACAACA	21781482
308942	cg23091758	55782376	AAAAACAACAACCTAATAAAAACTAACCCAAACA TCCAAAAATCCACCA	30707421
309640	cg23744638	35611335	CACAAAACTTTAAAAACACTCTAACCCCTACTACT CACCCATACAACA	14763344
314613	cg00748589	60757479	AACTCTACTCAAAAACTCAAAAACCATCTACTA AACCAACAAAACTC	NA
315365	cg01528542	50719302	TCCCACCTTACAACTTTACCAAAATTTAATCATCAC TAACAAAAATTAAC	NA
330233	cg18473521	28617472	ATTACCCATCTCCTCCTTAATCCAATTCAAATTAT ACTAACCCCAAAATC	NA
331413	cg19722847	25621361	CAACAATCTCTTAAACAACAAACCAATTACRAT ACCCATACAATTC	NA
350626	ch.13.39564907R	61804317	GCTACTTCCCTAAAATTAACAACCACTTATTACCCTC TCACCTAAACTACT	NA
352716	cg03032497	24810422	TATCTAACTCAACCCCTTTAAATATTTCTCCAAATA AAATTATAAATTC	NA
367851	cg03399905	52669494	CCCATACTCRACCTTCTAAAAAATACCCCAAAACAC TAACAATAATAAAC	NA
368628	cg04875128	66707375	AACTATAACTCTACACACACAAACAACAATTCCTAC TACAAATACATAACA	34724474
371123	cg09651136	69636474	CACTACACCCCAAACTTTAAAACTTTAAATAACA ATATAATCCAACATC	NA

图 9D

图 9D

ID	名称	地址_A_ID	地址_B_ID	等位基因_A_核苷酸序列	地址_B_ID
377332	cg21296230	29625433	NA	CRCRAAAATAAATTATAAAAACCATCRRAAATC CTTCTACTAAAACC	NA
388382	cg04416734	71765505	24794497	TCAAAATCAAACTCCAAATAAAAACCCAAAACCCCT ACCCTCTAAAACA	24794497
392814	cg07082267	49808346	NA	AAAAAAAACCATAAAATCTCCRTTTCACAATA AACACACAAAACCC	NA
406094	cg02867102	24773488	NA	AATAACTAAAAATCTATCTAAATCCAAAACRCTA ACTTTAACCTTCTC	NA
410441	cg06874016	71689361	NA	AAAAACAAAAACRAACACAAITATAATAATAAA CTTACTCTACAACC	NA
418324	cg14692377	11791474	74735323	AACTACAAAAAACTAATCCAAAACCTAAAACAAA CAACTAACCTCACA	74735323
424470	cg21139312	71738392	36705403	AAAAACCCCAAACTATAATTTCCAAAACAAA TACAAAACAACACA	36705403
434803	cg19283806	62709369	12681323	AATTTCTTTAAACAATCCCAACAAAATAACAAC CAAAAAAAATACA	12681323
449879	cg14556683	39707455	23657395	AAAAACCAAACTCCAGATAAAACACACAACAAA CTTCAACAACAACA	23657395
465513	cg07547549	72801360	NA	AACTCACTCCATTAAAATACTCCRAACRGTATCCA AAATACTAAAATAC	NA
479076	cg05442902	45650355	NA	CTCTCACTTATACCTTTAAITTTTACATACTCT AATCTTACATAC	NA
480022	cg08415592	18687383	NA	AACAAATCTTTCTCTTAAAACCAACAACCCCCC AACCCCTAAAATAC	NA

图9E

单位基因_核苷序列	Infinium_设计_类型	下一个_碱基	颜色_通道
AAAATAAAAACATCTCAAACCTCGGCTTACGAAA ACCGATTGAAAACCG	I	T	Red
CGAAACTACGACACCTTACGACGAAAACCAAAAT TTAAATCTAGCAACG	I	A	Red
	II		
	II		
	II		
	II		
	II		
	II		
	II		
	II		
	II		
	II		
ACGCGAAAACCTTTAAAAAACGACTCAACGA CTTAACCTTACCGTACG	I	C	Grn
CTTAAAAACACAATAATATCGAAAACGTCGCCT CCGAGTAACTCTCG	I	A	Red
TCTGAAAACCTTAACGACTTACCGCTAAAACCC GCAATACAAACAAACG	I	C	Grn
AAAACACAATAATATCGAAAACGTCGCCTCCG ACGTAACCTCTCGAACG	I	C	Grn
	II		

图 9A-1

图9A-1

等位基因ID_核苷酸序列	Infinium_设计_类型	下一个_碱基	颜色_通道
	II		
	II		
	II		
GTATAACGCCAAATAAAAAATAATTCTCAAAA ACCGTCGACACCGGACG	I	C	Gm
	II		
GCTATCTCTCAAAAACCGCCAAAATACTAAAAAA AACGACAACACGGAACG	I	C	Gm
	II		
	II		
	II		
	II		
	II		
	II		
	II		
	II		
TCTAAACACGGAATAATAATTCTAAACTCCGT AAAGTTAAACCGCCG	I	C	Gm
	II		

图 9B-1

图9B-1

定位基因组_探针序列	Infinium_设计_类型	下一个_碱基	颜色_通道
	II		
	II		
ACTAAACACGAAACGTACAAGGCCCAACC	I	A	Red
TCGTATAATAAAACCG			
ATCTCCGGTTCCTTCTTCTACCGACGACCAAA	I	A	Red
TAATAATAACAAAACG			
	II		
AATTCAAAACCTTAAACCCCAACGCTCTACAAA	I	A	Red
CAAAAACGAAACGGG			
	II		
ATAACGCTTAAACAACAATAACAACCTCCTC	I	C	Grn
CGAAACGCTAATCCG			
TCTTCTACCGACGAAACCAATAATAACAATAA	I	A	Red
CGAAACTCCCAATCG			
	II		
CAAAAACCTAATTTCTACAATGCTAAAACCTAC	I	C	Grn
AACGACGCTACTCCCG			
	II		
	II		
TTTCCACAGTACTCTGCTACTTTATTACCTA	I	A	Red
AACCAAAAAAACG			
ACACACCTCGAACGCTTAAACGAAATAGCT	I	A	Red
AAAAAACCAAAAATTCG			

图 9C-1

图 9C-1

每个基因E _n 的序列	Infinium 探针类型	下一个碱基	颜色/通道
TAATCAGTATCCTACCGCTACACTAAAAACTTC CGACTAACACAACG	I	A	Red
	II		
ATCTACTAACACACACCCGCAATCTAACACTCAT CTATATCCACAACG	I	A	Red
AAAAACGACGCGCTAATAAAAACTAACCGGAC AATCCGAAAATCCACCG	I	C	Gm
CAGGAACTTTAAAAACACTTAACCCCTACTA CTCACCCATACAACG	I	A	Red
	II		
	II		
	II		
	II		
	II		
	II		
	II		
AACTATACTCTACGACGACAACGACGATTCT ACTACGAATACGTAACG	I	C	Gm
	II		

图 9D-1

图9D-1

等位基因_探针序列	Infinium_设计_类型	下一个_碱基	颜色_通道
TCGAAATCAACTCGATATAAAACCCAAACCC CTACCCCTATAAAAACG	II		
	I	A	Red
	II		
	II		
	II		
AACTAGCGAAAAAACTAATCCGAACTAAACA AACGAATCAACTCGCG	I	C	Grn
AAAACCGCCCGAACTCTATAATTTCCAAAACA ATACGAAACGACACG	I	T	Red
AAATTCCTTTAAACAATCCCGCAAAAATAACA ACCAAAAAAATACG	I	A	Red
GAAAAACACAAACTCCACATAAAAAACGCGCAAC AACTTCAACGACAAACG	I	C	Grn
	III		
	II		
	II		

图 9E-1

正向_序列	基因组_构建	CHR	MAPINFO
GGGTATTTTCGGGGTGAAGGCACTCAGGCTCGGTTACGGGGACCGGTTTCGGGAGAC[CG]TGG	37	1	117665053
AGCCGAGGTGTCGAATGGAGGCTTACTTCGGCCGGCCAGGGAGCCACCTCCT	37	1	236557682
CCCAGAGACTTTCATCTAGAGGTTTGAAGTCTGCGCAGACACCCAGGAGCATCTT[CG]CAATC	37	1	207997020
TGTTGCTTCTCCATGGCAACTCCAGAGAATTAAGAAGCCAAACTCAACA TCGC	37	1	169556022
ACGTGGGGAAAGGGGTTAGCCATCAAGTCTGAAGCCGTCGGACACCCATCG[C]CT	37	1	17338766
GCGAGACCAAACTTGTCCCGGTAAGGTGCGCACTCCGAAATGTTCCAGAA	37	1	26855765
TTCA TAGCCGACAAAGCTGTATCCCTCCA TTTCCACTGCGCAACACCGGAAGCAGT[CG]TCCGT	37	1	28241577
TACCACTGACCTGAGGCTGCTGGTCCCAAGCTCACACTTGGAGAACCTCTGT	37	1	97202260
AA TGCCTTTCACAGAGAACTGCTGCGAGGATCACACAAGAAATGCTGTCAACTGGG[CG]TGGT	37	2	106015739
GCCGATGCTGTAAATCCCACTACTCCGAGACTAAGCCAGGAGAA TCGCTTGAAC	37	2	230452311
GGTCTGATGGTGAACAACCTGATGCTCCATC TTGGCTGGGCACTGAGATCGCC[CG]GAGA	37	2	39187539
TCAAGTGTAGCTTCAGGGGGGTAGAAATAGAGGATAGGGATCTTAGGGC	37	2	106015767
CTAGCTCACAGCACCGGTGAGTGTCTTTTACATAGGATCACACTCTCT[CG]TGAAT	37	2	131513927
GCCCAATGGTGCCTTCTCCCACTGGGCTTGAGAAACCTACGCTCT	37	2	106015771
AGTTTGCCTCCAGGAACTGAGGCACAAGGCAGCAATGATTA CTGAGGGTCTTGCTC[CG]CTCC	36	2	30561970
TCTAGGTGAGGAGCCTATCCAGGGCTCCAGTCTGAAAGCCTAGAGGCGAGGGC			
CTTTTGTTCACAGGCTCTTTCTTGTGCTCCGCTTTCGGGTCAGTAGTAT[CG]GGAGC			
GTGCTCCGGCTGCGGCTCTCGGGCGAGTTTCGGACGAGGCTTGGGGCCG			
CTACAGTCCCGCAGGGCTTAAAGGGAGCCAGTGCACACAGATTGGTTAACCGGCCA[CG]TGT			
CCAGTCTTCTAAAAATCCCAAGACTGGACAGCAATGCCCTTAGGTGATG			
AGAGGCTAGTTCGCGGCGGCGAGCAGCGCTTACAAACTGGAGGCGGGCGAGG[CG]CA			
CGGCAAGCCAAAGCCCTGAGCCCTCTCTAAAGATTCGCGGTGCGCCGCGCCAGCC			
TGCTCCGGTCTTGGAGCACAGTAGTTATCGGAGGCTGCTCCCGGGTGGGCT[CG]GGG			
CGAGTTTCGGAGGCTTGGGCGGCTGGGCGGCTGGCAGGGTCTGCCCCACCGGGATCTC			
CTGACGTTGCTGAGCGGCTGCGCTCCAGCGGGTCCCGAGCCTGTGGGGCCG[CG]CCT			
GCTGCACTGGGGCCCCCAGCGGTAAGTTCGCAAGGCCCCGAGAGGCTGCGTGGT			
CTCCGGTCTTGGGAGCACAGTAGTTATCGGGAGCGTCCCTCCGGGTGGGCTTCGGG[CG]CGA			
GTTTCGACAGGCTGGGCGGTTGGCAGGGTCTGCCACCGCGGGATCTGCC			
ACCGCTATCTGGTAGAGTAGAAGGAGGATGGCTTGAAGATCAATTTGGAGATTA[CA]TCCA			
TAGGACTTCTGATGATCACATGTGGCAGAGGAGGAGCCAGTCTCAAAATGGCC			

图 9A-2

图 9A-2

正向序列	基因组_构建	CHR	MAPINFO
TCAGAGGTTGCGTTTCAGCACCTGGGTCAGCGCTCCAGGGTCAGCCAGGGATAGA[CG]CCAT	37	3	187387650
TGTGGCTTCGTGCCAGACAACACTACCCAGATTTCGTAATTTCTTGAATGCCAAGT			
CGCTGTGGCCCGCCGAGCGGAACGGCCGGGAGAGAGCGGTCGCCGGGAAACCAAGTGC[CG]CCC	37	3	53080440
TGGCCAGCCCGATCCAGCTGCGCTCACCTCGGGTTGTAGACAGAGCGCCGGGG			
CCAGCAAGTGGCTTGTATCGTTTCCCAATGCCCGAGCCTGTTTCTTCCGCAAGTAGAG[CG]GGTCA	37	3	52008487
GATGTGCCAACTCTGCAGAGTAGCAATAAGCAGTAACGCCACGCTCTGCACA			
AATCCGATGGCACCGTGGCTGGGGAGAGGCTGGGCTGGCGGGGACGAGGGGAAG[CG]CC	37	3	160167977
GGTGGCGAGCGCTTCTGAGGAATTATCTTTACTTTGGCGCCACACGSGGGCGGGGCT			
TCGGAGCTGAGGGACCCGAAAAGCACAAAACCTTTAGAGGACTGAGCATCCCTTA[CG]TCCA	37	3	187387866
AACCAATGGGCGAGGCAAGGCTTAGGGAGGGCTGGAGAAATCCGGGAGAGCTCGA			
GCCCGAGAGGATCCAGGGAAAGCAGAAAGGGGTTAAGGACCTTGGACAGAGCCCGTCCGCG[CG]CT	37	4	147558239
CGTTGTCGCGCTTCCCGAGCACTTGGCGGCTCTGAGGACAGCGTCCCATCTTG			
GGGAGGTTGGGAGAGGGAAGAGGATGAGGATGAGAAATACCTATTAGATAAATGTACAACATT[CG]GGT	37	4	84255411
GACAGTTACACTAAGAGCCAGATGCCACCAATGCCAATATACCCATGGAAACAGAA			
TTGTATTCAGCCCAAAGCTACTGGAAATGTCAGTGGCAGCTCCCTCTGCCCCTCC[CG]TTGCTA	37	5	139013317
TGGCAGCCATGCTCTGTGTGTGTAATAGGTAAGCCAGGCTCCAGGTTAGGACCT			
CAGGAGTGGGTGCAGCCACATCCCAAGCTGACAGGGGGGCACTTGCCTAAGTCTT[CG]GCG	37	5	140419819
TGCTGCCITCCACAACACTTCTCAGCTTGTCTGATTTGAAAGGCTTAGTAA			
ATTCAGGGGAAATCAGAGGGTGGAGAGGGCGTGGTGGCTGGAGATGCCCTGGGAAACAGAA[CG]GCG	37	6	31105711
TGAGGGACTCCATTATCTGTACTTCCCGGGTGGGTTAGGTTAGGCTTGGCTCTCCCTG			
GGAGAGCAAGTCAAGAAATACGGTGAAGGAGTCCCTCCAAAGTTGTTAGGTTCTTCCG[CG]CCGG	37	6	25652602
TGCTGTCTTGTGTCAACACCAAGGAGCTCCCGGGAAACCACTTGGGGGG			
TCACCACTTTGGCAGTCTAGATCGTCTGGTCTTACTGTGCATACAGTTTACT[CG]TCTCAG	37	6	30295466
GTGAGGGCCCACTTAATTTGTAAGAGACTGAGGAAAGGTTAGGATCACCACAA			
TGTTTACTAAGTGAAGTCTACTATACATAGTAAATACGCAGTCACTTATAAA[CG]ACAAA	37	6	147528521
ATCTGACAGGCTAGATCTGAATCTGTTAATTTCTGAGAGGTTGTTCTGAAATG			
CCCGCGCTCCCTTCCCGCCGGCGGGAATTCAGGTTCCAGCCGGCGGTTTCGGG[CG]GCG	37	6	11044877
GCTCAAAGTCCAGAGCCCGGAAATACCCAGCCGCTGGCCAGATGCGGAGCCGCT			
CTTGGCTCGAATGCCAGTGAATCTCTCACTTGGAGACATGCTTCCCTT[CG]AGCTGC	37	6	10510789
TCCAGTCCAGGTGAGGAAATGCTATGATGGCGTGAAATTTCCCTTGGT			

图 9B-2

图 9B-2

正向_序列	基因结构	CHR	MARINFO
TGCGCAGGGCCGACGAGGCGGACAGCCACAGTGGCCGAGGACAGGTTGGCGGGG[CG]CC GCTGTCGGGTGGCCAAACTTCTCAAGACACACCTTGCACTCGAGCAGGCTGATCTC	37	6	1812719
AAATTAAGACTAATTCAGAAATTTCAAGTGTAGTAAACAACACTGCTATCTCAACAACATA[CG]ATA AAATGAAACCACTGGTGCCTAAGTCTTCTCACTCAAACTCTGCTGTGA	37	7	131008743
CGTGGCTGGCCCAAAAGCCGCGGGGCTCGGACTACAGCGAAGCCGCGGGGCT[CG]GC CTCATAACAGGCTGGGCGCTGCACGCCCCGCTGCTTACGCCGCGGCTCCCG	37	7	15643108
AGAAATTACTGCTCCGCGTCTTCTTCTCGCGGAGCCAGGTAATGGTAACAGAG[CG]AAACT CCCAAGTCGAACTCTGGTGTGACAGAGCCTGCCCCCTCCCGCGAGCCGCG	37	7	130419159
CCGCTAAGTCATGTTAACAGCCTCAGAAATATCTTCTCGGCTTCTTCTTCTG[CG]GAGC CAGGTAATGGTAACAGAGCAAACTCCCAAGTCGGAACCTCTGGGTTGCGCAG	37	7	130419133
CCCGAGGCGGACGCCAGAGGGGCGCCCCCTCTCGCCGCTCGGGGCGGAGG[CG]CGC TCCGCCCTTGGCTGCAGAGGCTGGGGTTAAAGTCTCTGAACCCATGCAAGGCTG	37	7	152591698
TGGAAACGCGGCCCCCGGCTAAGTCAATGTTAACAGCCTCAGAAATATCTTGTCT[CG]GTT TTTCTTGCAGGAGCCAGGTAATGGTAACAGAGCAAACTCCCAAGTCGGA	37	7	130419116
GAAAGGCAATGGTGGCGCTTGGCAGCAGGTTGACAGACTCTCCGCGGCGCTGAT[CG]CGG CGGGGCGGGGCTGCCCTAGGCCCCCTCCAGAAACCCAGAGGCTGCTGGTG	37	7	130418324
CTCCGCTTCTTCTGCGGAGCCAGGTAATGGTAACAGAGCGAACTCCCAAGT[CG]GAACT TCTGGGTTGACAGCAGCCTGCCCCCTCCCGCAGCCGCGCGCTGGTCCG	37	7	130419173
CAAACAGGGGAGCCCCATTAAGGTTTGTACTAGGATCAATCGAAACTT[CG]GATTC CTAGTTATAGAGTTGAATCAACCAACCAACACTCCAGAACTCTGACATTAGG	37	8	82644012
CTGCCCCCTCGAAAGGGCTGATTTCTACAGTCCCTAGGACCTGCGGGGCGCTGCT[CG]CGGG GCTCGGCGCGCTGCATGCTCCATTAGTCCCTAGAGGCGAGCCCTCTCTGCG	37	8	57360613
GCAACACAGAGTCTTACCCATGCTTAAAGGAGCTGTCATGACTTCCCTTAATA[CG]GTTGT GAGGTCTCAGTGTGAGATAGCAAAAGTCTCAAATAATAGGACAGACG	37	10	98062687
TCCTAAGCCTCTGAGCTGGCTGGCCACTTCCGGGTTGAGCCTCCAGGGAGAT[CG]ACCA CACAGGCAACCCAGGAGCAAGTCTTGAATGCGGCTTCTCCGACCTTGCAGG	37	10	49673534
GAGTCAGTGCACAGATTAGAGGAGGAGAAAGAAAGAGACTCCCAACTCGGCC[CG]CCC TTCCCTGGCCAGGCAAAAGCAGCGAGGAGCCAGTGTGAAAGCAGTGCAGAT	37	11	133797911
GCCACACACGACACACTCTGGGCGCTTGGACGGGGTGGCTGGGGAGCCAGAAAT[CG]GAG CGAGCCGGGCGGAGAGCCCGCTCGGAGCCCGGAGCCCGGCTGCAACCCCT	37	11	12696758

图 9C-2

图 9C-2

正向序列	基因组_构建	CHR	MAPINFO
GGTGGTTGTCGGGGGTTGAAATGTGAAGAACAATCGCGGGTCTTCTGCTGAGGC[CG]CGG ACACCGTGAACCTGGCTGCTGCTGGGTCTGCAAGGAAACGTAGGAAAAAAGTTGTCAAG	37	15	33010536
ACTGTCTGCTTCAGATCAAGCTCCGATGAGGACCCAGGGCCCTGCCCCCTGCGGGAG[CG]GCCA GCCCCAGGGCCCATGTGCCCTCTCCCTGAAGACCCTTCCCCACGCCACTGGAA	37	16	30075192
GCTCTCATGTGAGAAGGACCATAGGAACTCCCGTTTCAAGTGGGGCCACCAAGGC[CG]ACAA TGGGTCAGGCTGCCAAGGGTGGAGCCAGATGCAAGGGGACCTCAGAGCCTGC	37	16	85429035
CTCCTGGAGTGGGTGCTCCTGGGATGCTTCAAGTTTAGACACCGGGTTACGGCAGCTG[CG]AGGA AGGCTAAAGCCAGCGTCTGGATTCAAGACAGACCTTTAGCCATTAATCCACTAA	37	17	62398693
CAGCCTTCAGGAGCTGACAGTCTCTTTGGGGCTCAGGAGGTTGGGACACACCCAG[CG]GCCT GCAGAGTAAGCTTATTACCCACAACCTGTGCCCGCTTTGTCTTCTAAGGTGCACAC	37	17	40177415
ACCCACTGCCCAGGCTGCGGGGAGGCTGCCCCGGCTGGGCAAGCGGGCTGGCCTCG[CG]CCC TCAGGCACCCGGCGGCTGGCTGTGCGGAGGGGCGCGCGCCGTATTTGTA	37	17	28562685
TTGGAGCTGCCAAAGCCGCCCGGGCTCTGTGTTTCCAGAGCAGATGCGGAGCGGCA[CG]TCTT CGTGGCTTGTCTCCAGTGGCCACACGACTCAGCCTCTCTGCCCCGCTGGGCGC	37	17	55663225
TCCGTAGTATTGTTCTGGCTTTGAGCGTGTGAGGGAGGGAAATGTTGCACTCATCC[CG]CATCC TTTTTTGGCTGCTATCTTGGGGGATTGTTCAAGGAGAAATCCATCCTGACTGG	37	18	66389420
ATCCAGCCACCAGAACACCCAGGCTCCACATGAAGCGCGCAGCTTCAGCGACAGG[CG]CGA CGGCGCACAGCGCGGTCACCCAGCTCCGGCATGTGCCCGCTCCGGGACCAAC	37	19	15342982
TTGCAGCTGGAGCTCAGCTCCAATGGAAATGCTCCGGGCTGTCCAAGGTGCTGGAA[CG]CGGC GCCCGGGGCGAGAGCTGCGGGCCGGGATATGCTGCCCCAGGCTTCCGGCTGA	37	20	44658225
GCCAGGTACCCCTCACTCTGCTCTTAGTATTCTGATGCTGCTGCTTGGAT[CG]CTGCTCC CTGCACAGAACTCCATCCCATCTTGTCTGCTTGTGCAACTTCAGAAAT	37	22	21369010
AGTATGTCAGTGGCAGTCTTCTCTTGAAGCCACAGACCCAGCCCTGAGGATG[CG]JAGGC AGGTGGTTGGATGAGAGGGATCTGGATGCTGGTCTCAGGCTGCTCTCTTAAGGG	37	22	36648973

图 9E-2

源序列	染色体_36	配位的_36	探针_SNP#	探针_SNP#	探针_SNP#_10
CGGTCTCCGAAACCGGTCCCGTAACCGAGCCCTGAG					
ATGCCCTCACCC	1	117466576 R			rs78180333
TTTCATCTAGAAGGTTTGACTCTGCCAGACACCAGC					
GAGCACTCTCG	1	234624305 R			
CGGGACTGGGCACCTTACGGCGGGACCAAGATTG					
GGTCTGCGAGCG	1	206063643 F	rs61821292		
CGTCCGTACCACTGACCTGAGGCTGCTGGGTCCA					
AGTCCACACTGG	1	167822646 F	rs9332483		
CGCCAGTTGACAAGCATTTCTTTGTGATCTCGCA					
GCAGTCTCTGT	1	17711353 R			
CGGGGATCTCAGTGGCCCGCCAGGCAAGATGGGCATC					
AGCAGTTGTACCA	1	26728352 R	rs59827434		
CACGGGTGGAGTTGCTTGTCTTTTACATAGGAGGT					
CACATCTCTCG	1	28114164 R			
TAGGCTTTCAGACTGGAGCCCTGGAATAGGCTCCTC					
ACCTAGAGAGCG	2	96565987 F			rs13020129
AGGGCTCTTTCTTGTGTCCTCCGGTCTTGGGAGC					
ACAGTAGTATCG	2	105382171 R			
CGTGGCCTGGTTAACCAATCTGTTGCACTGGCTCCCTT					
TTAAGGGCCCTG	2	230160555 R			
ACGGGAACTTTGAGAGAGCGGCTCAGCGGCTT					
GGCCTTGCCTGCG	2	39041043 F			
CTTGGGAGCACAGTAGTTATCGGGAGCGTCCCTCCG					
GCTGGGCTCTCG	2	105382199 R			
TCTGGGGCCCTTGGGACTTACCGCTGGGGGCCCGCA					
GTGAGCAGGGCG	2	131230397 F			
CGCCGAGAGCCCAAGCGGGAGGCGAGCTCCCGAT					
AACTACTGTGCTCC	2	105382203 R			
CATCCATAGGACTTCTGATGATCACAATGTGGCAGAG					
GAAGAGCCAGTCC	2	30415474 F			

图 9A-3

图 9A-3

源序列	染色体_36	配位的_36	链	探针	探针_SNP_s	探针_SNP_s_10
GTTTCAGCACCTGGGTCAGCGCTTCCAGGGTCAGCA CCAGGATAGACS	3	188870344	R			
CGCCCTGGCCCGCCGATCCAGCCTGGCCTCACC TCGGTTGTAGAC	3	53055480	F			
CGGGTCAGATGTTGCCAACCTCTGCAGGTAGCAATA AGCAGTAAACGCC	3	51983527	F			
CGCCGTGGCCGACGGCTCTGAGGAATTAICTTTTA CTTGGCCACAC	3	161650671	F			
GGACCCGAAAAGCACCAAACTCTTTAGAAGGACT GAGCATCCCTTACC	3	188870560	R			
GCTGCTCTCAGGAGCCGACAGGTGCTGGGGAAGGC GGCAGCAACGAGCG	4	147777689	F		rs76832102	
CGGGTGACAGTTACACTAAGAGCCAGATGCCACCAC AATGCAATATACC	4	84474435	F			
GGAGCCTGGTTCACCTATTCACACAGAGACATGGC TGCCATAGCAACG	5	138993501	F			
GCAGCCACATCCAAGGCTGACAGGGGGGCACTC TGCCAAGTCTGCG	5	140400003	R			
CGGCTGAGGGGACTCCATTATCTGTACTCTCCCGGG GTGGGTCTAGGTC	6	31213690	F	rs35474509		
AAGAAATACGGTGAAGGAGTCTCCCAAAGTTGTCT AGGTCCTCCGCG	6	25760581	R			
TGCCAGTCTAGATCGTCTGCTTACTTACTGTGATA CAGTTCTACTCG	6	30403445	R			
CGACAAAATCGACCAGGCTAGATCTGAATCTGTTAA TTTCTGAGAGGGT	6	147570214	F			
CGGGCGCTCAACGTCCACGGAGCCCGAAGAAATACCC ACCCGTGCCAGA	6	11152863	F			
CGAGCTGCTCCAGTCCAGGTGAGGAATAAAATGCTAT GATGGCCGTGAAAA	6	10618775	F			

图 9B-3

图 9B-3

源序列	染色体_36	位置_36	碱	探针_SNP	探针_SNP_s_10
CTCGAGTGC AAGGTGCTTTGAGAAAGTTGGCCACC	6	18230698	F	rs28940575	
GGCAGACGGCG					
CGTATGTTTGAGATAGCAGTGTGTTACTATCACTTG	7	130659283	R		rs76007347
AAAAATCTGAAT					
CGGCCCTCACTACACGAGGCTGGGCGCTGCAGGCC	7	156125869	F		rs74424274
CCCGTCTCAGC					
CGCTCTGTTACCATACCTGGCTCGCCGGGAGAAGAA	7	130069699	R		
AGAACGGGAGAC					
TGTTTAAACAGCCTCAGAAATTTATCTGCTCCGCTTC	7	130069673	R		
TTTCTCTGCCG					
CGCGCTCGCCCTTGGCTGCAGAGCGCTGGGGTTT	7	152222631	F		
AAAGTCTGAACC					
GCCCCGGCTAAGTATGTTTAAAGGCTCAGAAAT	7	130069656	R		
TATCTGTCTCCG					
GTGGGCTTGGCAGCAGGTTGACAGACCTCTCCG	7	130068864	R		
GGGGCCTGATCCG					
TCTTTCGGCGCAGCCAGGTAATGTTAACAGAGCG	7	130069713	R		
AAACTCCCAATCG					
CAGCCCATTTAAGGTTTTGATACACTGAGGATCAT	8	82806567	R		
CAGAAACTTCG					
CAAAGGCTGATTTACAGTCGCTAGGACCTGCAGC	8	57523167	R		
GGCGTCTCCG					
CGGTTGAGGTGCTCAGTGTGTTGGAGATAGCAA	10	98052677	F		
AGTCTCAAATAAT					
CGACCACAGGCACCAGGAGCAAGTCTTTGAA	10	49343540	F	rs11101335	
ATCGGCTTCTCC					
TTTCCACAGTGGCTCCTCGCTGCTTTGTTGCTGGGC	11	133303121	F		
CAGGAGGGGCG					
ACACACCTCGGGCCTTGGACGGGTTGCGCTGGG	11	12653334	R		
GAGCCAGAAATTCG					

图 9C-3

图 9C-3

源序列	染色体_36	配位的_36	雄	探针_SNP	探针_SNP_10
TGGTCACGTGCTCCGCTGCACCTGAGGGCTCCGG CTGGCAGGACG	11	65840273	F		
CGGCAATCTGGCACTCATCTGTGCCACACAGCGGTG CGGATCATCTCC	11	65840117	F		
ATCTGTGGCACACACCAGCAATCTGGCACTCATCTG TGCCACACAGCG	11	65840149	R		
GGAGGGGGGGGCTGGTGGGACTGACCCGGCAG TCCGAGAATCCACG	11	8982343	R		
CACGAAGCTTGGGGAGCACTTAGCCCTGCTACTC ACCCATGCAAGCG	11	10280478	R		
GCTCTACCTAAGGAGCTCAGGGCCATCTGCTGAAC CAACAGGGCTCG	12	11544753	R		
CGTTAACTCTGCTAGTAGTACCAAACTGGTAAAG ATTGTAAGTGGG	12	79992363	F		
TTACCCATTCTGGCTGTAATCCAGTCAATTGTGCT AACCCAGAGTCG	12	52734532	F		
AACAAGTCTATCTGGAACAGGCCAGTTGGGATGCC CATACAGTCTCG	12	30740381	F		
CAGTAGTCCAGGTGAGAGTAAATGGCCCTGCCTAA CCAGGGAAGCAG	13	39564907	F		
ATCTAACTAACCCCTTTAGATAATCTCCAGGTGGAA TTATTGGATTCC	14	6017980	R		
CCATGCTGGCTTCTGGAAAGATGCCACAGACACTG GCAATAATGGACG	15	77363115	R		
CGCCAGTACCGCAGCAGAAACCGCTGCTGCTCG CAGAGCTACAGCC	15	29563187	R		
ACTGCACCCAGACAATTTAGAACTTTAAAATACAATAT AGTCCAGCATCG	15	70312066	F		

图 9D-3

图 9D-3

源序列	染色体_36	位置的_36	链	探针_SNP's	探针_SNP's_10
GCGGGGTGAATTGTGAAGAACCATCGGGGTCTCT TCCTGTAGGCCG	15	30797828	R		
TCGAGATCAAGCTCCGATGAGGCCCCAGGGCCCTG CCTCTGGGAGCG	16	29982693	R		rs36029842
CGGGCTTGGTGTGCCACCTGTGAACGGGAGATTC CTATGCTCTCT	16	83985536	R		
ATGGCTAAAAGGTCTGTGAATCCAGGACGCTGGCT TTAGCCTCTCG	17	59752425	F	rs71377739	
CGGCTGCAGAGTAAGCTTATCCACAACTGTGCC CGCTTGTGCTTC	17	37430941	F	rs1128719	
GGCTGCGGGGGAGGCTGGTCCCGGGCTGGGCAAG CGGGCTGGCCTCGCG	17	25586811	R		
CGTCCGCTCCGCATCTGCTCTGGAACCACAGAGC CCGGGGCGGCTT	17	53018224	R		
GATTTCTCTTGAACAATCCCGCAAGATAGCAGCC AAAAAGGATGCG	18	64540400	F		
GAGAACCCAGGCTCCACATGAAGGCCGCGCAGCAGC TTCAGCGACAGGCG	19	15203982	R		
GCTCAGCTCCATTGGAATGCTCCGGGCGCTGTCCAAG GTGCTGGAATGCG	20	44091632	R		
TCTCACTGTGCTCTTAGTTATCTTGCATGCTCTGGT CTTTGCATCG	22	19699010	R		
CGCATCTCAGGGCTGGGGTCTGCTGTGGTCTCAAG GAGAAAGACTGC	22	34978919	R		

图 9E-3

基因组_Loci	甲基127_Loci	UCSC_RefGene_名称	UCSC_RefGene_增加
NA	NA	TRIM45;TRIM45	NM_025188;NM_001145635
NA	TRUE	EDARADD;EDARADD;EDARADD	NM_080738;NM_145861;NM_145861
NA	NA		NM_000130
NA	NA	F5	NM_022089;NM_001141974;NM_001141973
NA	NA	ATP13A2;ATP13A2;ATP13A2	NM_002953
NA	NA	RP56KA1	NM_002946
NA	NA	RPA2	NM_212481
NA	NA	ARID5A	NM_001039492;NM_001450;NM_201557;NM_201555
NA	NA	FHL2;FHL2;FHL2;FHL2	NM_139072
NA	NA	DNER	NR_028386;NM_001145451
NA	NA	LOC375196;LOC100271715	NM_001039492;NM_001450;NM_201557;NM_201555
NA	NA	FHL2;FHL2;FHL2	NM_152698;NM_001105194;NM_001105195;NM_001105193;NM_001105195
NA	NA	FAM123C;FAM123C;FAM123C;FAM123C;FAM123C;FAM123C	NM_001039492;NM_001450;NM_201557;NM_201555
NA	NA	FHL2;FHL2;FHL2;FHL2	
NA	NA		

图 9A-4

图 9A-4

随机的_Loci	甲基127_Loci	UCSC_RefGene_名称	UCSC_RefGene_编号
NA	NA	SST	NM_001048
NA	NA	SFMBT1;SFMBT1;SFMBT1	NM_001005159;NM_016329;NM_001005158
NA	TRUE	ABHD14B;ABHD14B;ABHD14B;ABHD14A;ABHD14B;ABHD14B	NM_001146314;NR_027476;NM_032750;NM_015407;NM_032750;NM_001146314
NA	NA	TRIMS9	NM_173084
NA	NA	SST	NM_001048
NA	NA		
NA	NA	HPSE;HPSE;HPSE	NM_001098540;NM_001166498;NM_006665
NA	NA		
NA	NA		
NA	NA	PSORS1C1;PSORS1C2	NM_014068;NM_014069
NA	TRUE	SCGN;SCGN	NM_006998;NM_006998
NA	NA	HCG18;TRIM39;TRIM39;HCG18;TRIM39	NR_024053;NM_172016;NM_021253;NR_024052;NM_172016
NA	NA	STXBPS;STXBPS	NM_001127715;NM_139244
NA	NA	ELOVL2	NM_017770
NA	NA		

图 9B-4

图 9B-4

融合的_Loci	甲基127_Loci	UCSC_RefGene_名称	UCSC_RefGene_增加
NA	TRUE	NHLRC1	NM_198586
NA	NA	MKLN1	NM_001145354
NA	NA	C7orf13;RNF32	NR_026865;NM_030936
NA	NA	KLF14	NM_138693
NA	NA	KLF14	NM_138693
NA	NA		
NA	NA	KLF14	NM_138693
NA	NA	KLF14	NM_138693
NA	NA	KLF14	NM_138693
NA	NA	CHMP4C	NM_152284
NA	NA	PENK	NM_001135690
NA	NA	DNTT;DNTT	NM_004088;NM_001017520
NA	NA	ARHGAP22	NM_021226
NA	NA	IGSF9B	NM_014987
NA	NA	TEAD1	NM_021961

图 9C-4

图 9C-4

随机的_Loci	甲基127_Loci	UCSC_RefGene_名称	UCSC_RefGene_增加
NA	NA	CD248	NM_020404
NA	NA	CD248	NM_020404
NA	NA	CD248	NM_020404
NA	NA	NRIP3	NM_020645
NA	NA		
NA	NA		
NA	NA		
NA	NA	HOXC4;HOXC4	NM_153633;NM_014620
NA	TRUE	IPO8	NM_006390
NA	NA		
NA	NA		
NA	NA	ANKRD34C	NM_001146341
NA	NA	OTUD7A	NM_130901
NA	NA	PKM2;PKM2;PKM2	NM_182470;NM_182471;NM_002654

图 9D-4

图9D-4

随机的_Loci	甲基12T_Loci	UCSC_RefGene_名称	UCSC_RefGene_增加
NA	TRUE	GREM1	NM_013372
NA	NA	ALDOA;ALDOA;ALDOA	NM_001127617;NM_184043;NM_000034
NA	NA		
NA	NA		
NA	NA	NKIRAS2;NKIRAS2;NKIRAS2;NKIRAS2	NM_001144927;NM_001144928;NM_017595;NM_001144929;NM_001001349
NA	NA	SLC6A4;SLC6A4	NM_001045;NM_001045
NA	NA	MSI2;MSI2	NM_138962;NM_170721
NA	NA	CCDC102B	NM_001093729
NA	NA	EPHX3;EPHX3	NM_024794;NM_001142886
NA	NA	SLC12A5;SLC12A5	NM_020708;NM_001134771
NA	TRUE	MGC16703;P2RX6;P2RX6	NR_003608;NM_005446;NM_001159554
NA	NA	APOL1;APOL1;APOL1;APOL1	NM_001136541;NM_003661;NM_001136540;NM_145343

图9E-4

UCSC_RefGene_Group	UCSC_Cpg_岛名称	相关_to_UCSC_Cpg_岛
TSS1500;TSS1500	chr1:117664180-117665148	Island
TSS1500;5'UTR;1stExon	chr1:236558459-236559336	N_Shore
TSS1500		
TSS1500;TSS1500;TSS1500	chr1:17337829-17338590	S_Shore
TSS1500	chr1:26856191-26856684	N_Shore
TSS1500	chr1:28240584-28241535	S_Shore
TSS1500	chr2:97202313-97203787	N_Shore
TSS200;TSS200;5'UTR;TSS200	chr2:106014878-106015884	Island
Body		
TSS200;Body	chr2:39186777-39187968	Island
TSS200;TSS200;5'UTR;TSS200	chr2:106014878-106015884	Island
5'UTR;5'UTR;1stExon;1stExon;5'UTR;5'UTR	chr2:131513363-131514183	Island
TSS200;TSS200;5'UTR;TSS200	chr2:106014878-106015884	Island

图 9A-5

图 9A-5

UCSC RefGene 组	UCSC CpG 岛 名称	相关_to_UCSC CpG 岛
Body	chr3:187387914-187388176	N_Shore
TSS1500;TSS1500;TSS1500	chr3:53078956-53081101	Island
1stExon;Body;5'UTR;TSS1500;1stExon;5'UTR	chr3:52008943-52009339	N_Shore
TSS1500	chr3:160167184-160168200	Island
Body	chr3:187387914-187388176	N_Shore
	chr4:147558231-147558583	Island
Body;Body;Body	chr4:84255726-84256399	N_Shore
	chr5:139017133-139017668	N_Shelf
Body;3'UTR		
1stExon;5'UTR	chr6:25652380-25652709	Island
TSS1500;5'UTR;5'UTR;TSS1500;1stExon	chr6:30294169-30295071	S_Shore
Body;Body	chr6:147525374-147525848	S_Shelf
TSS1500	chr6:11043913-11045206	Island

图 9B-5

图 9B-5

UCSC RefName 组	UCSC CpG 岛 名称	相关 to UCSC CpG 岛
1stExon	chr6:18122250-18122994	Island
Body	chr7:131012460-131013190	N_Shelf
Body;TSS1500	chr7:156432433-156433670	Island
TSS1500	chr7:130417912-130419378	Island
TSS1500	chr7:130417912-130419378	Island
TSS1500	chr7:152591458-152591706	Island
TSS1500	chr7:130417912-130419378	Island
1stExon	chr7:130417912-130419378	Island
TSS1500	chr7:130417912-130419378	Island
TSS1500	chr8:82644603-82644849	N_Shore
TSS1500	chr8:57360585-57360815	Island
TSS1500;TSS1500		
Body	chr10:49674243-49674776	N_Shore
Body	chr11:133800684-133800931	N_Shelf
5'UTR	chr11:12695414-12696981	Island

图 9C-5

图 9C-5

UCSC_RefGene_Group	UCSC_Cpg_岛_名称	相关_to_UCSC_Cpg_岛
1stExon	chr11:66083572-66083782	Island
1stExon	chr11:66083572-66083782	N_Shore
1stExon	chr11:66083572-66083782	Island
TSS200	chr11:9025095-9026315	Island
	chr11:10324353-10324828	N_Shore
	chr12:11653232-11653775	Island
	chr12:81471569-81472119	N_Shelf
Body;Body	chr12:54447744-54448091	S_Shore
TSS1500	chr12:30848264-30849016	S_Shore
	chr14:61108954-61109786	N_Shore
5'UTR	chr15:79576059-79576270	Island
Body	chr15:31775540-31776988	Island
TSS1500;TSS1500;TSS1500	chr15:72522131-72524238	S_Shore

图 9D-5

图 9D-5

UCSC_RefGene_Group	UCSC_CpG_岛_名称	相关_to_UCSC_CpG_岛
5'UTR	chr15:33009530-33011696	Island
TSS1500;TSS1500;5'UTR	chr16:30076310-30077872	N_Shore
3'UTR;3'UTR;3'UTR;3'UTR		
1stExon;5'UTR	chr17:28562387-28563186	Island
Body;Body		
5'UTR		
1stExon;Body	chr19:15342626-15343181	Island
Body;Body	chr20:44657463-44659243	Island
TSS1500;TSS1500;TSS1500	chr22:71368197-71368771	S_Shore
TSS200;TSS200;TSS200;TSS200		

图9E-5

体型	DMR	神器	PMR_岛	牌照_特征_名称
		NA	1:117465578-117466781	1:117663907-117665512
		NA		
	CDMR	TRUE	1:206063625-206063801	
		NA		1:169555452-169556050
		TRUE		1:17336920-17338827
		NA		1:26855689-26857507
		NA	1:28113187-28114165	1:28240552-28241702
		NA		2:97202196-97202767
		NA	2:105381311-105382817	2:106014507-106016259
		TRUE		2:230451331-230452578
		TRUE	2:39040222-39041697	2:39187021-39187940
		NA	2:105381311-105382817	2:106014507-106016259
	DMR	NA	2:131229834-131230653	2:131513688-131513993
		NA	2:105381311-105382817	2:106014507-106016259
		NA		

图 9A-6

图9A-6

体型	DMR	增强	HM_岛	调控特征名称
high-CpG:53054999-53055997	RDMR	TRUE	3:188870246-188870359	
		NA	3:53053856-53056190	
		TRUE	3:51982751-51985402	3:52007848-52009460
		NA	3:161649892-161650878	3:160166409-160168278
	RDMR	TRUE	3:188870501-188870889	
		NA	4:147777501-147778016	4:147557996-147558356
		NA		4:84255006-84256489
		TRUE		5:139013288-139013559
	RDMR	NA	5:140400003-140400154	
		NA		6:31104495-31106407
	DMR	NA	6:25760360-25760750	6:25652510-25652746
		NA		
		NA		6:147528521-147528685
		NA	6:11151611-11153237	6:11044102-11044892
		TRUE		6:10510346-10511316

图 9B-6

图 9B-6

体型	DMR	增强	IMN_号	调控_特征_名称
		TRUE	6:18230230-18231229	6:18122473-18123542
		NA		7:131008672-131009115
	DMR	NA	7:156125195-156126707	7:156432754-156434135
		NA	7:130068467-130069793	7:130418325-130419878
		NA	7:130068467-130069793	7:130418325-130419878
		NA	7:152222028-152222744	7:152590901-152592150
		NA	7:130068467-130069793	7:130418325-130419878
	DMR	NA	7:130068467-130069793	
		NA	7:130068467-130069793	7:130418325-130419878
		NA		
		NA	8:57522950-57523369	8:57360377-57362115
		NA		
		TRUE	11:133303046-133303162	11:133797084-133799070
		NA	11:12651991-12653557	11:12695339-12696865

图 9C-6

图 9C-6

体型	DMR	增强	IMEI_号	调控_特征_名称
		NA	11:65840060-65841164	
	RDMR	NA	11:65840060-65841164	
		NA	11:65840060-65841164	
	DMR	NA	11:8981699-8983012	
		NA		
		NA	12:11544500-1154529	12:11653353-11654101
		NA		
		NA	12:52734084-52734533	12:54447856-54448358
		NA	12:30739425-30740382	
		NA		
	RDMR	TRUE	14:60177929-60179820	
	RDMR	NA	15:77363046-77363443	
		NA	15:29562601-29564280	
		NA		

图 9D-6

图 9D-6

体型	DMR	增强	HMM_高	调控_特征_名称
		NA	15:30796823-30799072	
		NA		
		TRUE		
		NA	17:59752099-59752445	17:62397669-62399390
		NA		17:40176771-40177869
	DMR	NA	17:25586344-25587312	17:28562266-28563419
		NA	17:53018213-53018286	
	DMR	NA		18:66388995-66389733
	DMR	NA	19:15203635-15204238	19:15341951-15343455
		NA	20:44090882-44092713	20:44657985-44658436
		TRUE		22:21368080-21369261
		NA		22:36648819-36649508

图9E-6

调控特征_组	DHS	范围_START	范围_END	范围_CB	范围_ID
启动子_相关性	NA	117665053	117665176	NC_000001.10	
	NA	236557682	236557805	NC_000001.10	
	TRUE	207997020	207997143	NC_000001.10	
未分类_细胞_类型_特定的	NA	169556022	169556145	NC_000001.10	
启动子_相关性	NA	17338766	17338889	NC_000001.10	
启动子_相关性	NA	26855765	26855888	NC_000001.10	
非基因_相关性	NA	28241577	28241700	NC_000001.10	
启动子_相关性	NA	97202260	97202383	NC_000002.11	
未分类_细胞_类型_特定的	TRUE	106015739	106015862	NC_000002.11	
未分类_细胞_类型_特定的	NA	230452311	230452434	NC_000002.11	
未分类	TRUE	39187539	39187662	NC_000002.11	
未分类_细胞_类型_特定的	TRUE	106015767	106015890	NC_000002.11	
未分类_细胞_类型_特定的	TRUE	131513927	131514050	NC_000002.11	
未分类_细胞_类型_特定的	TRUE	106015771	106015894	NC_000002.11	
	NA	30561970	30562093	NC_000002.11	

图 9A-7

图 9A-7

调控_特征_组	DHS	范围_START	范围_END	范围_CB	范围_ID
	NA	187387650	187387773	NC_000003.11	
	NA	53080440	53080563	NC_000003.11	
启动子_相关性	NA	52008487	52008610	NC_000003.11	
启动子_相关性	NA	160167977	160168100	NC_000003.11	
	NA	187387866	187387989	NC_000003.11	
未分类_细胞_类型_特定的	TRUE	147558239	147558362	NC_000004.11	
未分类_细胞_类型_特定的	NA	84255411	84255534	NC_000004.11	
未分类	NA	139013317	139013440	NC_000005.9	
	NA	140419819	140419942	NC_000005.9	
未分类	NA	31105711	31105834	NC_000006.11	
未分类_细胞_类型_特定的	TRUE	25652602	25652725	NC_000006.11	
	NA	30295466	30295589	NC_000006.11	
未分类_细胞_类型_特定的	NA	147528521	147528644	NC_000006.11	
未分类_细胞_类型_特定的	TRUE	11044877	11045000	NC_000006.11	
未分类	NA	10510789	10510912	NC_000006.11	

图 9B-7

图 9B-7

调控_特征_组	DHS	范围_START	范围_END	范围_CB	范围_ID
启动子_相关性	TRUE	18122719	18122842	NC_000006.11	
未分类_细胞_类型_特定的	NA	131008743	131008866	NC_000007.13	
未分类	NA	156433108	156433231	NC_000007.13	
未分类	TRUE	130419159	130419282	NC_000007.13	
未分类	TRUE	130419133	130419256	NC_000007.13	
未分类	TRUE	152591698	152591821	NC_000007.13	
未分类	TRUE	130419116	130419239	NC_000007.13	
未分类	TRUE	130418324	130418447	NC_000007.13	
未分类	TRUE	130419173	130419296	NC_000007.13	
	NA	82644012	82644135	NC_000008.10	
未分类_细胞_类型_特定的	TRUE	57360613	57360736	NC_000008.10	
	NA	98062687	98062810	NC_000010.10	
	NA	49673534	49673657	NC_000010.10	
启动子_相关性	NA	133797911	133798034	NC_000011.9	
未分类_细胞_类型_特定的	NA	12696758	12696881	NC_000011.9	

图 9C-7

图 9C-7

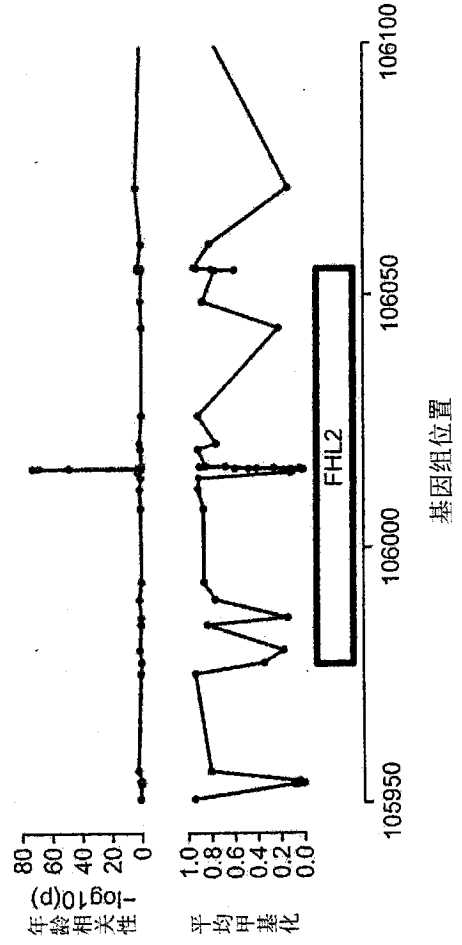
测控_特征_组	DHS	范围_START	范围_END	范围_GB	范围_ID
	NA	66083697	66083820	NC_000011.9	
	NA	66083541	66083664	NC_000011.9	
	NA	66083573	66083696	NC_000011.9	
	TRUE	9025767	9025890	NC_000011.9	
	NA	10323902	10324025	NC_000011.9	
未分类	TRUE	11653486	11653609	NC_000012.11	
	NA	81468232	81468355	NC_000012.11	
未分类	TRUE	54448265	54448388	NC_000012.11	
	NA	30849114	30849237	NC_000012.11	
	NA	40666907	40667030	NC_000013.10	
	NA	61108227	61108350	NC_000014.8	
	TRUE	79576060	79576183	NC_000015.9	
	NA	31775895	31776018	NC_000015.9	
	NA	72525012	72525135	NC_000015.9	

图 9D-7

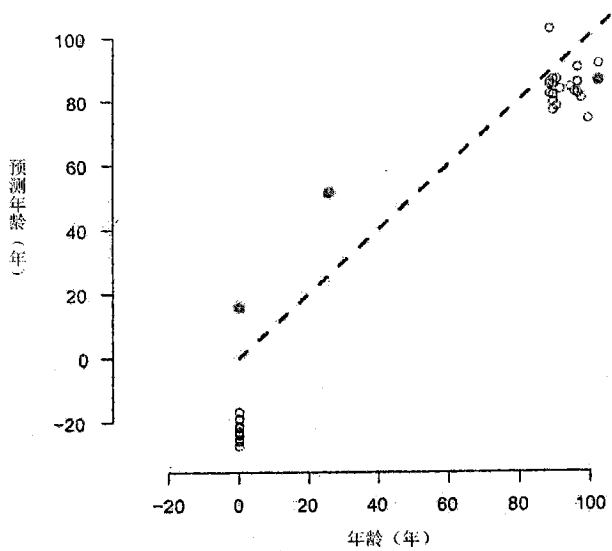
图 9D-7

调控_特征_组	DHS	范围_START	范围_END	范围_GB	范围_ID
	TRUE	33010536	33010659	NC_000015.9	
	NA	30075192	30075315	NC_000016.9	
	NA	85429035	85429158	NC_000016.9	
未分类	TRUE	62398693	62398816	NC_000017.10	
Gene_相关性	NA	40177415	40177538	NC_000017.10	
未分类_细胞_类型_特定的	TRUE	28562685	28562808	NC_000017.10	
	NA	55663225	55663348	NC_000017.10	
未分类_细胞_类型_特定的	NA	66389420	66389543	NC_000018.9	
未分类_细胞_类型_特定的	NA	15342982	15343105	NC_000019.9	
未分类_细胞_类型_特定的	TRUE	44658225	44658348	NC_000020.10	
启动子_相关性	NA	21369010	21369133	NC_000022.10	
未分类	TRUE	36648973	36649096	NC_000022.10	

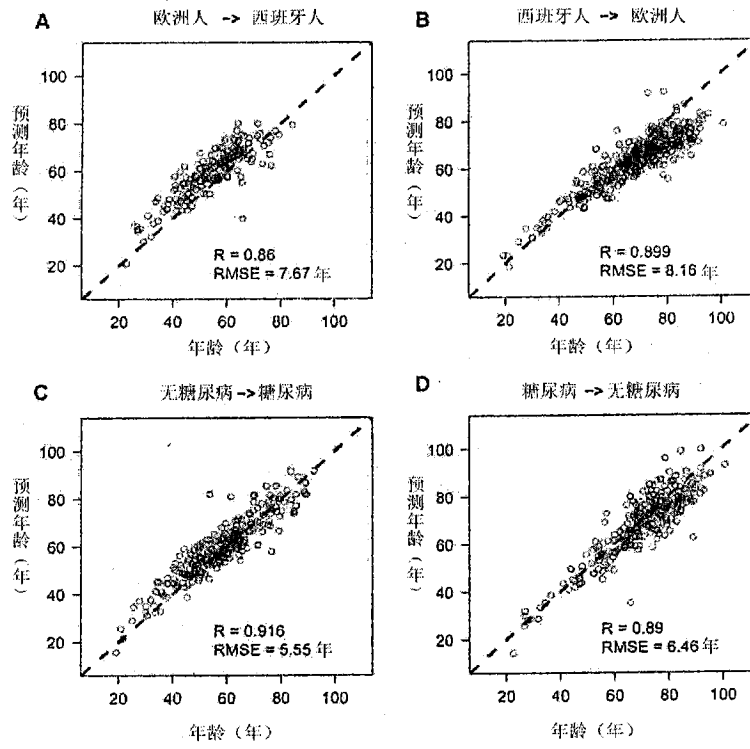
图9E-7



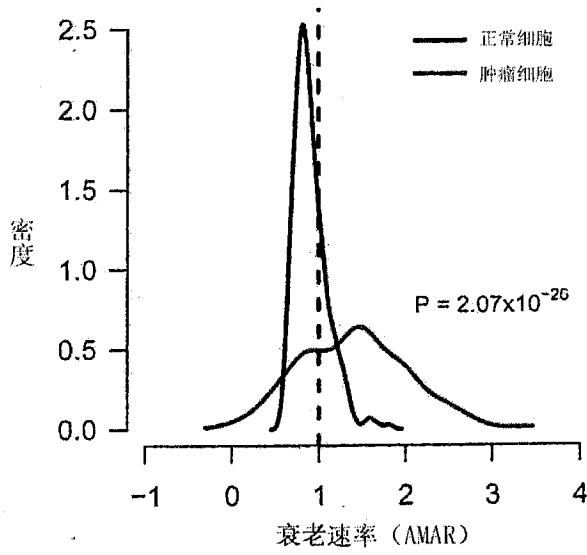
图S1



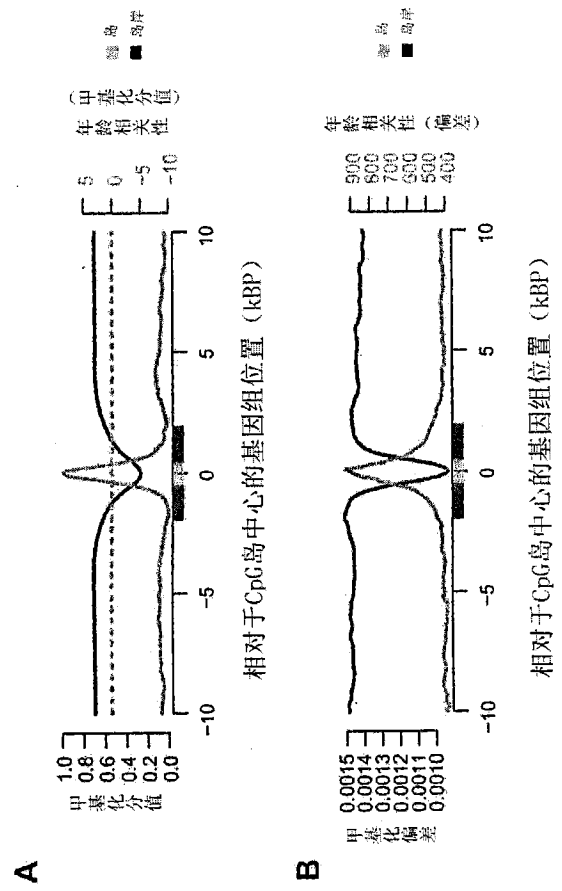
图S2



图S3



图S4



图S5

专利名称(译)	用于预测年龄和鉴别诱发或者抑制早衰试剂的方法		
公开(公告)号	CN105745333A	公开(公告)日	2016-07-06
申请号	CN201380070066.1	申请日	2013-11-12
[标]申请(专利权)人(译)	加利福尼亚大学董事会		
申请(专利权)人(译)	加利福尼亚大学董事会		
当前申请(专利权)人(译)	加利福尼亚大学董事会		
[标]发明人	康张 格里高利汉恩 崔伊伊德克 贾斯廷吉尼 史蒂芬H弗兰德		
发明人	康·张 格里高利·汉恩 崔伊·伊德克 贾斯廷·吉尼 史蒂芬·H·弗兰德		
IPC分类号	C12Q1/68 C12N15/11 G01N33/48 G01N33/53		
CPC分类号	C12Q1/6883 C12Q1/6881 C12Q2600/118 C12Q2600/136 C12Q2600/154 C12Q2600/16 G01N33/5308 G01N33/57484 G01N2500/00		
优先权	61/724528 2012-11-09 US		
外部链接	Espacenet SIPO		

摘要(译)

本发明提供了基于受试者的表观基因组预测受试者年龄的方法。

