

doi: 10.3969/j.issn.1674-1242.2023.04.005

DNA 甲基化年龄预测研究进展

张黎¹, 田若婷¹, 王子文^{1,2}, 王陈偲¹, 朱德旭¹, 王晗²

(1. 长春工业大学计算机科学与工程学院, 吉林长春 130102;

2. 东北师范大学信息科学与技术学院, 吉林长春 130117)

【摘要】衰老是生物体多个器官、组织和细胞功能下降的一个过程, 与衰老相关的疾病会在衰老过程中逐渐出现, 同时伴随着表观遗传学机制的变化。相关领域的研究也表明, 表观遗传学机制的变化过程与生物年龄的增长密切相关。随着全球老龄化现象的加剧, 抗衰老逐渐成为一个备受关注的研究方向, 然而衰老过程的表观遗传学表征、机理和定量分析方式仍然需要深入研究。该文对当前基于 DNA 甲基化的衰老与年龄预测的领域研究进展进行梳理, 介绍表观遗传时钟、生物时钟和 DNA 甲基化的基本原理, 以及相应表观遗传时钟的典型研究方法和有效样本资源, 并深入讨论该领域研究存在的挑战及发展趋势, 旨在为研究者提供全面的参考, 促进抗衰老相关临床实践和大健康应用的发展。

【关键词】衰老; 年龄预测; DNA 甲基化

【中图分类号】R339.38

【文献标志码】A

文章编号: 1674-1242 (2023) 04-0357-07

Advances in DNA Methylation Age Prediction Research

ZHANG Li¹, TIAN Ruoting¹, WANG Ziwen^{1,2}, WANG Chensi¹, ZHU Dexu¹, WANG Han²

(1. School of Computer Science and Engineering, Changchun University of Technology, Changchun, Jilin 130102, China;

2. School of Northeast Normal University Information Science and Technology, Changchun, Jilin 130117, China)

【Abstract】 Aging is a process in which the functions of multiple organs, tissues and cells of an organism decline, and diseases related to aging will gradually appear during the aging process, accompanied by changes in epigenetic mechanisms, and studies in related fields have also shown that the process of changes in epigenetic mechanisms is closely related to biological ageing. With the intensification of global aging phenomenon, anti-aging has gradually become a research direction of great interest, however, the epigenetic characterization, mechanism and quantitative analysis of the aging process still need to be studied in depth. In this paper, we review the current research progress in the field of aging and age prediction based on DNA methylation, introduce the basic principles of epigenetic clock, biological clock and DNA methylation, as well as the typical research methods and effective sample resources of the corresponding epigenetic clocks, and discuss in-depth the challenges of the research in this field as well as the trend of development, with the aim of providing comprehensive references for the researchers to facilitate the anti-aging related clinical practice and the development of large health applications.

【Key words】 Aging; Age Prediction; DNA Methylation

收稿日期: 2023-12-08。

基金项目: 国家自然科学基金 (62372099)、吉林省科技发展计划 (20230201090GX、20230401092YY、20210101175JC)、吉林省预算内基本建设资金 (2022C043-2)。

作者简介: 张黎 (1979—), 女, 吉林省长春市人, 硕士研究生导师, 从事退行性疾病研究。

田若婷 (2000—), 女, 吉林省梅河口市人, 硕士研究生, 从事退行性疾病研究。

通信作者: 王晗, 男, 副教授, 博士研究生导师, 电话 (Tel.): 13086869997, 邮箱 (E-mail): wangh101@nenu.edu.cn。

0 引言

近年来,多项研究指出,表观遗传机制受环境、生活方式、遗传和疾病等因素的影响会发生变化,进而影响人体的衰老过程^[1-3]。其中,DNA甲基化作为一种典型的表观遗传修饰,与衰老相关疾病存在显著的相关性^[4]。由于DNA甲基化在一定程度上受基因组控制,基因组中近1/3的甲基化位点与年龄有关^[5],且大量文献表明基因组区域的甲基化值会随着年龄的增长而发生变化^[6]。CpG位点是DNA分子中的一种特定序列,在基因组中广泛存在,它们的主要特点是容易发生DNA甲基化修饰。人体基因组中的大量CpG位点会随着时间的推移而发生DNA甲基化的动态变化,表现为甲基化的增加或减少^[7]。利用这些特定的CpG位点作为特征,在构建表观遗传时钟时预测个体的年龄,可以为衰老相关研究提供更准确和更可靠的支持。多项研究使用多个CpG位点作为特征来预测个体的年龄,其中正则线性回归模型Elastic Net是早期预测个体年龄的典型方法^[8]。Fan等^[9]应用机器学习方法,如逐步回归、支持向量回归、随机森林回归等,构建表观遗传时钟。而近些年来也有多种深度学习方法被作为构建表观遗传时钟的方法,并在一定程度上提高了预测精度。例如,Galkin等^[10]应用人工神经网络构建表观遗传时钟DeepMAge;Levy等^[11]开发了一种DNAm深度学习方法MethylNet,可以研究细胞差异,掌握癌症亚型的高阶信息等。

1 基于DNA甲基化的生物年龄预测器

1.1 表观遗传时钟

表观遗传时钟利用特定CpG位点的DNA甲基化水平构建预测模型,从而预测生物年龄^[12]。特定CpG位点主要选取自不同基因组区域与年龄高度相关的位点,利用CpG位点的DNA甲基化水平的变化(特异性CpG位点的甲基化水平在不同DNA来源的组织或细胞中呈现出差异,包括高甲基化和低甲基化^[13])进行年龄预测。基于DNA甲基化的表观遗传时钟是采用随年龄变化的CpG位点创建的,其中随年龄变化的CpG位点主要选自训练数据中和年龄呈单调递增关系的CpG位点,简而言之,就是与年龄相关的高甲基化或低甲基化的关键CpG位点^[12]。研究表明,DNA甲基化,特别是5-甲基胞嘧啶(5MC),可以作为一种准确的生物标志物来估计生物年龄^[14]。它不仅可以在

生物体衰老过程中起到一定的作用与影响,而且可以预测衰老状态及与年龄相关的变化。

近10年来的研究指出,人体基因组中大量CpG位点的DNA甲基化会随着时间的推移表现出增加或减少的趋势,由此可以选择这些与年龄高度相关的CpG位点来测量个体的衰老速度^[7,15]。最初,表观遗传时钟是从整个基因组中挑选少量与年龄高度相关的CpG位点创建的。2011年,Koch和Wagner^[3]使用来自真皮、表皮、宫颈涂片、T细胞和单核细胞的19个数据集训练了一个表观遗传时钟,该时钟使用5个CpG位点来预测多种细胞类型的年龄。2014年,Weidner等^[15]基于3个CpG位点作为生物标志物进行血液年龄预测。随着技术的进步,如今可以利用全基因组的DNA甲基化数据来创建更加准确的表观遗传时钟。这些表观遗传时钟不仅可以根据不同的组织类型和物种进行优化,还可以预测生物属性(死亡、功能衰退时间和疾病风险等)。Levine等^[16]使用513个CpG位点创建了一种新型表观遗传生物时钟DNAmPhenoAge,它在预测各种衰老结果(包括全因死亡率、癌症、健康寿命、身体机能和阿尔茨海默病)方面效果远远优于以前的措施。Lu等^[17]创建了新的表观遗传时钟GrimAge,并证明GrimAge对死亡时间、冠心病患病时间、癌症患病时间、绝经年龄等的预测能力优于现有的其他表观遗传时钟。

创建表观遗传时钟之后,选择CpG位点的 β 值(甲基化百分比)在模型中进行加权计算,不断学习,完成年龄预测。通过建立和优化年龄预测模型,捕捉不同尺度和位置下的不同年龄特征,进而提高模型预测的精度、可靠性及应用性。

1.2 生物时钟

生物年龄是多个因素协同作用的结果,如基因、环境、生活方式等因素都会影响人体的衰老过程^[2]。一对双胞胎拥有相同的时间年龄,在不同的环境和生活方式下,最终可能拥有不同的生物年龄^[18]。也就是说,拥有相同时间年龄的个体在受到与年龄相关的疾病、功能衰退、死亡等因素的影响后,可能拥有不同的生物年龄。

现阶段的所有表观遗传时钟在实际年龄预测方面均有较高的精度和准确性,但是时间时钟和生物时钟存在一定的区别。尽管两种时钟都与时间相关,但是

时间时钟仅包括甲基化模式下与年龄强烈相关的 CpG 位点；生物时钟的 CpG 位点更加广泛，它包括与生物属性（如健康状态、疾病风险、生理功能衰退等）相关的 CpG 位点。因此，在捕获 CpG 位点时，选择相同时间年龄下生物属性和生理状态不同的 CpG 位点^[16]比选择随时间变化的 CpG 位点更重要。时间时钟主要用来评估个体的生物年龄与实际年龄之间的差异；生物时钟不仅可以预测死亡风险，还可以预测多种疾病状态。时间时钟主要应用在年龄预测领域；生物时钟主要应用在人类衰老研究领域。

尽管生物年龄存在复杂性，不可能以单一组织的单一定量值来表示，但是相信未来真正的生物时钟会出现，可以强有力地预测疾病和衰老。真正的生物时钟不仅可以精准地预测年龄，还可以捕获特定的生物衰老特征、功能退化情况，并能预测疾病。如今需要在人类和动物群体中进行更多的研究，以选择最佳 CpG 位点和训练生物时钟的最佳方法，从而进一步了解组织和疾病特异性的重要性，并测试候选生物时钟的预测价值。

2 基于 DNA 甲基化的表观遗传时钟

2.1 DNA 甲基化检测技术

在建立基于 DNA 甲基化的表观遗传时钟时，DNA 甲基化检测是一个必不可少的步骤。因为随着年龄的增长，不同来源组织或细胞中的 DNA 甲基化水平会发生变化，这种变化不仅有序，还可以预测。DNA 甲基化测序的主要作用是定量和定位不同组织与细胞类型在基因组范围内的甲基化水平，并探究其与年龄之间的关系。

Benjamin 等^[19]以斑马鱼为研究对象，开发了一种多重 PCR 检测方法。该方法不仅花费较少，还可以有效评估 DNA 甲基化水平，从而评估年龄的准确性。Vetter 等^[20]应用甲基化敏感的单核苷酸引物延伸法和 Infinium 甲基化 Epic 阵列方法评估 DNA 甲基化，基于两种不同评估方法的 DNA 甲基化年龄估计值之间存在 3.1 年的差异。Knight 等^[21]采用新生儿脐血和血斑样本，应用弹性网络回归模型计算胎龄，其中所使用的甲基化数据是使用 Illumina Infinium Humanmethylation 27K BeadChip 或 Illumina Infinium Humanmethylation 450K BeadChip 平台生成的。Illumina Infinium Human Methylation 27K BeadChip 或 Illumina Infinium Human

Methylation 450K BeadChip 芯片技术是常用的评估 DNA 甲基化水平的技术，并被广泛应用于开发和验证基于 DNA 甲基化的表观遗传时钟。

值得注意的是，不同的 DNA 甲基化测序技术的覆盖性、准确性等存在差异。因此，在构建基于 DNA 甲基化的表观遗传时钟时，需要结合多种元素，如样本数量、样本年龄分布特征等，以选择合适的 DNA 甲基化测序技术。各种测序技术都有各自的优点、缺点及适用场景。

2.2 特征选择方法

利用 DNA 甲基化检测技术分析数据中 CpG 位点的甲基化状态，计算 β 值（该 CpG 位点的甲基化胞嘧啶的比例），采用特征选择方法帮助筛选与年龄相关且重要的 DNA 甲基化位点，从而构建出更加精确和可靠的表观遗传时钟。尽管特征选择方法通常可能不是最好的方法，但是在面对一些情况时，它是一个重要的步骤。在面对大量的候选特征进行筛选时，特征选择方法可以发现并区分存在显著差异的 CpG 位点，从而减少噪声和冗余信息。当数据有限，观测量少、预测量多时，会出现模型在训练集上过度拟合而在测试集上预测效果较差的情况。这种情况可以应用特征选择方法查看对输出变量有重要作用的特征，以此仅将其作为输入^[22]。也可以应用特征选择方法对多个 DNA 甲基化位点进行组合与优化，达到提升信息丰富度与复杂度的目地，从而进一步提高模型的预测能力和泛化能力。

Lee 等^[23]使用弹性网络回归模型自动选择了 558 个 CpG 位点作为 RPC 模型（稳健胎盘时钟），并对 CpG 位点的 DNAm 水平的 GA（因变量）进行回归。Vidal-Bralo 等^[24]通过前向逐步线性回归的方法进行特征选择，从而选择数据集中信息最丰富的位点，每一步都应用 MS-SNuPE 来分析纳入模型的 CpG 位点是否可行，最终共 8 个 CpG 位点被纳入模型。Wei 等^[25]提出了基于重采样的特征选择方法 FESTWO，该方法在只有 70 个特征的测试数据集上使 RMSE 降低了 8% 以上，从而改善了年龄回归模型的回归性能。Li 等^[26]应用新的特征选择方法和组合方法，包括新适应的神经网络、遗传算法和“链式”组合，使用较少的 CpG 位点构建了一个精确的表观遗传时钟。

2.3 模型构建

构建表观遗传时钟的基本原理是，在包含 DNA

甲基化数据和年龄信息的数据集上使用机器学习方法或深度学习方法来学习,从而以较高的精度预测年龄,并在不同年龄段的数据集上验证训练模型的准确性^[26]。大约10年前,大部分基于DNA甲基化的表观遗传时钟使用的是惩罚回归模型,如弹性网络、岭回归和Lasso回归^[7,27]。现在,研究人员不断应用新的方法构建了新的表观遗传时钟,并且具备高精度和高准确性。

Alfonso等^[28]通过结合贝叶斯神经网络对Horvath方法进行了改进,使用Elastic Net对数据降维,得到353个变量,并将它们作为输入,继续使用BNN降维,使用Levenberg-Marquardt算法和贝叶斯正则化训练神经网络,最终预测准确性达到0.94。Fan等^[9]利用不同的机器学习算法建立了基于AR-CPGS的年龄预测模型($r \geq 0.7$),如逐步回归(SR模型)、支持向量回归(SVR-EPS模型和SVR-NU模型)和随机森林回归(RFR模型)。SR模型、SVR-EPS模型、SVR-NU模型和RFR模型的中位数绝对偏差(MAD)值分别为2.97年、2.22年、2.19年和1.29年。Freire-Aradas等^[29]提出了一个年龄预测模型。该模型采用年龄平均分布在2~104岁的895名西班牙人的DNA样本作为数据集,使用位于7个基因组区域的7个CpG位点的甲基化水平,通过比较分位数回归、分位数回归神经网络和分位数回归支持向量机3种模型,对年龄预测的准确性进行了检验。最终,得到分位数回归神经网络模型和分位数回归支持向量机模型的MAE分别为 ± 3.36 岁和 ± 3.41 岁,预测准确性较高。Ambroa-Conde等^[30]建立了一个涵盖唾液和颊部细胞模型的年龄预测模型。该模型使用年龄分布在21~86岁的184个样本(唾液和颊细胞)作为数据集,采用7个CpG位点,使用多变量分位数回归进行年龄预测,最终MAE为 ± 3.66 岁。

2.4 有效样本资源

基于DNA甲基化的表观遗传时钟常使用GEO数据库和TCGA数据库中的DNA甲基化数据。GEO数据库收录并整理了全球范围内研究者上传的微阵列芯片、二代测序及其他形式的高通量基因组数据,同时存储了大量芯片数据,包括原始数据和处理后的数据。TCGA数据库收录了各种人类癌症(包括亚型肿瘤)的临床数据、基因组变异、mRNA表达、甲基化等数据,涵盖33种不同类型的癌症样本和相关临床资料,

包括大约10 000个肿瘤样本和相应的对照样本。

虽然在实验中使用的数据库大多为GEO数据库和TCGA数据库,但是也有其他数据库可供研究。EWAS Data Hub数据库有75 344个样本的DNA甲基化阵列数据,并采用有效的归一化方法来消除不同数据集之间的批量影响。该数据库能提供不同背景下的DNA甲基化图谱,涉及81种组织和细胞类型(包含25个脑部和25种血细胞类型),以及67种疾病(包含39种癌症)。MethBank数据库是一个十分全面的公共DNA甲基化数据库,该数据库收集和整合了来自各种物种(包括人类、小鼠、斑马鱼、草履虫和植物等)生物样本中的全基因组或基因组区域的DNA甲基化数据。Pubmeth数据库收集和整理了文献中与癌症相关的甲基化数据,并进行了人工校对和注释,是一个高质量癌症相关的甲基化基因数据库。

以上列举的数据库只是众多数据库中的一部分,未来还需探究更多的数据库,为研究做出贡献。

3 挑战与展望

3.1 模型优化问题

目前,在建立表观遗传时钟时,CpG位点的常用筛选标准是选择与年龄相关的位点,但是该标准存在一定的缺陷。虽然选择与年龄相关的CpG位点提高了预测年龄的精度,但是忽略了CpG位点并不一定只受年龄的影响,还受其他因素的影响,如环境暴露、疾病状态、生活方式等。这些因素会对DNA甲基化产生影响,使与年龄相关的CpG位点无法完全反映年龄本身。在选择CpG位点时,不仅要考虑与年龄相关度高的CpG位点,还要考虑与生物属性相关的CpG位点。但是,所考虑的生物属性越多,测量的CpG位点数量越多。如果没有优良的特征选择方法,会使关键CpG位点的选择变得困难,所以使用特征选择方法来构建精确和广泛性强的时钟。在特征方法的选择上,可以考虑单个算法或多个算法相结合等形式。Li等^[26]提出了应用神经网络算法和遗传算法,神经网络算法的优点是含有较多的参数,可以对多个参数进行优化。遗传算法由于参数优化途径广泛,所以在未来应用超特定参数调优方面拥有巨大的潜力。多个算法相结合的形式是将不同的特征选择方法相结合和组合。在选择和组合特征算法时,可以将不同的算法进行优势互补,实现更全面的特征选择,从而选出其他方法无法选出

的 CpG 位点。

基于 DNA 甲基化的表观遗传年龄预测模型有弹性网络、岭回归、深度学习、随机森林、支持向量机等。Horvath^[7]和 Hannum 等^[27]的表观遗传时钟采用弹性网络模型来识别与年龄相关度高的 CpG 位点,从而得到非常准确的年龄预测。但是弹性网络、岭回归等模型都存在处理特征向量较多的高维数据时能力有限的问题。弹性网络模型在处理高维数据时容易出现过拟合问题。岭回归模型在处理特征数量较多的高维数据时,会导致模型保守,达不到最佳预测效果。深度学习模型在处理大规模数据时可以通过引入更复杂的网络结构和提高特征提取能力等方式提高模型的预测能力。随机森林模型和支持向量机模型尽管可以处理高维数据,但是当遇到大规模数据时,存在训练时间较长和效率较低的问题。相比而言,深度学习模型通过神经网络端到端的训练或自动学习特征的复杂表示,可以处理大规模数据效率低问题。支持向量机模型在超参数的选择上较为敏感,但深度学习模型可以通过并行计算提高训练效率,具有更强的自适应能力,并能降低对超参数的依赖。深度学习模型不仅可以应用在年龄预测领域,也可以应用在癌症预测领域。

3.2 样本问题

研究人员在提高模型的精度与准确性方面耗费了大量的时间和精力。除了构建越来越精确的表观遗传时钟,样本问题也不容忽视。对于构建基于 DNA 甲基化的表观遗传时钟,采用横截面数据具有一定的弊端:首先,横截面数据无法揭示个体或种群的变化趋势,导致其不能完全捕捉到衰老过程中的动态性和个体差异;其次,横截面数据忽略了时间维度,可能导致无法准确预测个体或种群在未来表观遗传变化和衰老方面的状态。应用横截面数据将限制分析评估时钟变化的能力。因此,对大规模及多样化纵向样本进行研究是非常有价值的。在整个生命周期中,个体的差异具有多样性,包括生命起始的差异、特定生命转折点(青春期等)等,这些变化在分析和评估表观遗传时钟时都做出了一定的贡献,所以跟踪个体一生的纵向研究在评估疾病发展或预测疾病衰老方面具有巨大的优势。纵向研究可以跟踪个体或种群随时间的变化,提供更详细的时间解析度,从而更好地理解衰老过程及

表观遗传变化的动态性。通过纵向观察,可以观察到个体之间的差异,并探索这些差异与表观遗传变化之间的关联。这些有助于确定不同因素对衰老和表观遗传变化的影响。目前,大多数表观遗传时钟研究应用的是成年人的横截面数据,少数进行的是初步纵向研究。Horvath^[7]和 Hannum 等^[27]研究发现,表观遗传年龄随时间的增长速度较慢,并且在青少年时期的表观遗传时钟中有一个非线性(对数)模式^[31]。因此,Horvath 提出的表观遗传时钟在应用纵向数据时,在个体晚年时表现出渐近线的迹象,在 7 岁和 15~17 岁的人群中发现 CpG 位点的 DNA 甲基化程度不同^[32],由此可见,该模型在对个体衰老、疾病和 DNA 甲基化生物标记物方面的预测能力很强。

研究表观遗传时钟时,大规模生成 DNA 甲基化数据及多样化种群是有益的选择。许多表观遗传数据较多地选自欧洲人群,对亚洲人群的研究较少,使其代表性不足。人类基因组具有多样性,不同人群和族群之间的 CpG 位点存在差异,如日本冲绳地区的人均寿命长于大部分亚洲人群。所以未来的研究应当重视跨种群和跨族群的研究,考虑进行大规模跨种群和跨族群的比较研究,以提高年龄预测模型在不同人群中的适用性和准确性。

如今大部分高精度的年龄预测模型采用较少的 CpG 位点进行预测,使预测局限在小部分 CpG 位点上,无法全面覆盖所有与衰老、疾病及死亡相关的 CpG 位点。扩大样本的规模是有必要的,这样可以使样本包括更多更全面的 CpG 位点信息,这有助于寻找新的 CpG 位点并提高模型的泛化能力。在优化 CpG 位点的选择上,可以考虑使用基因组学和生物信息学的方法(数据挖掘技术、机器学习算法等),这样可以发现与年龄相关的特定 CpG 位点,从而进行更准确的年龄预测。

4 总结

本文对当前基于 DNA 甲基化的年龄预测研究进展进行了梳理,基于 DNA 甲基化的年龄预测模型在过去几年取得了显著的成果。通过 DNA 甲基化的变化,这些模型能够较为准确地预测个体的生物年龄。基于 DNA 甲基化的人类年龄预测研究对于理解人体衰老过程、提供个性化医疗和抗衰老疗法具有重要意义。然而,该领域仍存在一些挑战和问题,如目前的模型

仍然面临使用的 CpG 位点数量有限、种群样本不够多样、数据质量和可靠性不足、数据隐私和道德等问题。未来的研究应当重视跨种群的比较研究,以更好地理解不同种群之间的衰老差异。此外,研究人员还应探索多学科的合作,将年龄预测模型与其他生物标志物相结合,以提供更全面、更准确的衰老评估。总之,未来的研究应继续致力于克服当前模型的局限性,提高年龄预测的准确性和可靠性,同时深入探索衰老的分子机制,为相关健康管理和抗衰老研究提供更有力的支持。

参考文献

- [1] TSALENCHUK M, GENTLEMAN S M, MARZI S J. Linking environmental risk factors with epigenetic mechanisms in Parkinson's disease[J]. *npj Parkinson's Disease*, 2023, 9(1).
- [2] NOROOZI R, GHAFOURI-FARD S, PISAREK A, *et al.* DNA methylation-based age clocks: from age prediction to age reversion[J]. *Ageing Res Rev*, 2021, 68: 101314.
- [3] KOCH C M, WAGNER W. Epigenetic-aging-signature to determine age in different[J]. *Ageing*, 2011, 3: 1018-1027.
- [4] 高杰, 沈成, 黄新河. 衰老的表现遗传调控机制[J]. *中国生物化学与分子生物学报*, 2017, 33 (11): 1098-1104.
GAO Jie, SHEN Cheng, HUANG Xinhe. Epigenetic regulation mechanisms of aging[J]. *Chinese Journal of Biochemistry and Molecular Biology*, 2017, 33(11): 1098-1104.
- [5] JOHANSSON A, ENROTH S, GYLLENSTEN U. Continuous aging of the human DNA methylome throughout the human lifespan[J]. *PLoS One*, 2013, 8(6): e67378.
- [6] 孟航, 马开军, 董利民, 等. 基于 DNA 甲基化推断年龄的研究进展[J]. *法医学杂志*, 2019, 35 (5): 537-544.
MENG Hang, MA Kaijun, DONG Limin, *et al.* Research Progress on Age Estimation Based on DNA Methylation[J]. *Journal of Forensic Medicine*, 2019, 35(5): 537-544.
- [7] HORVATH S. DNA methylation age of human tissues and cell types[J]. *Genome Biology*, 2013.
- [8] LI A, KOCH Z, IDEKER T. Epigenetic aging: biological age prediction and informing a mechanistic theory of aging[J]. *Journal of Internal Medicine*, 2022, 292(5): 733-744.
- [9] FAN H, XIE Q, ZHANG Z, *et al.* Chronological age prediction: developmental evaluation of DNA methylation-based machine learning models[J]. *Front Bioeng Biotechnol*, 2021, 9: 819991.
- [10] GALKIN F, MAMOSHINA P, KOCHETOV K, *et al.* DeepMAge: a methylation aging clock developed with deep learning[J]. *Ageing Dis*, 2021, 12(5): 1252-1262.
- [11] LEVY J J, TITUS A J, PETERSEN C L, *et al.* MethylNet: an automated and modular deep learning approach for DNA methylation analysis[J]. *BMC Bioinformatics*, 2020, 21(1): 108.
- [12] SIMPSON D J, CHANDRA T. Epigenetic age prediction[J]. *Ageing Cell*, 2021, 20(9): e13452.
- [13] XU C, QU H, WANG G, *et al.* A novel strategy for forensic age prediction by DNA methylation and support vector regression model[J]. *Sci Rep*, 2015, 5: 17788.
- [14] JIANG S, GUO Y. Epigenetic clock: DNA methylation in aging[J]. *Stem Cells Int*, 2020, 2020: 1047896.
- [15] WEIDNER C, LIN Q, KOCH C, *et al.* Aging of blood can be tracked by DNA methylation changes at just three CpG sites[J]. *Genome Biology*, 2014, 15(2): R24.
- [16] LEVINE M E, LU A T, QUACH A, *et al.* An Epigenetic Biomarker of Aging for Lifespan and Healthspan[J]. *Ageing*, 2018, 10(4): 573-591.
- [17] LU A T, QUACH A, WILSON J G, *et al.* DNA methylation GrimAge strongly predicts lifespan and healthspan[J]. *Ageing*, 2019, 11: 303-327.
- [18] FIELD A E, ROBERTSON N A, WANG T, *et al.* DNA methylation clocks in aging: categories, causes, and consequences[J]. *Mol Cell*, 2018, 71(6): 882-895.
- [19] BENJAMIN M, DARREN K, LISA K, *et al.* A DNA methylation age predictor for zebrafish[J]. *Ageing Cell*, 2020, 12(24): 24817-24835.
- [20] VETTER V M, KALIES C H, SOMMERER Y, *et al.* Seven-CpG DNA methylation age determined by single nucleotide primer extension and illumina's Infinium methylation EPIC array provide highly comparable results[J]. *Front Genet*, 2021, 12: 759357.
- [21] KNIGHT A K, CRAIG J M, THEDA C, *et al.* An epigenetic clock for gestational age at birth based on blood methylation data[J]. *Genome Biology*, 2016, 17(1): 206.
- [22] MAYNE B T, LEEMAQZ S Y, SMITH A K, *et al.* Accelerated placental aging in early onset preeclampsia pregnancies identified by DNA methylation[J]. *Epigenomics*, 2017, 9(3): 279-289.
- [23] LEE Y, CHOUFANI S, WEKSBERG R, *et al.* Placental epigenetic clocks- estimating gestational age using placental DNA methylation levels[J]. *Ageing Cell*, 2019, 11(12): 4238-4253.
- [24] VIDAL-BRALO L, LOPEZ-GOLAN Y, GONZALEZ A. Simplified assay for epigenetic age estimation in whole blood of adults[J]. *Front Genet*, 2016, 7: 126.
- [25] WEI Z, DING S, DUAN M, *et al.* FeSTwo, a two-step feature

- selection algorithm based on feature engineering and sampling for the chronological age regression problem[J]. **Comput Biol Med**, 2020, 125: 104008.
- [26] LI A, MUELLER A, ENGLISH B, *et al.* Novel feature selection methods for construction of accurate epigenetic clocks[J]. **PLoS Comput Biol**, 2022, 18(8): e1009938.
- [27] HANNUM G, GUINNEY J, ZHAO L, *et al.* Genome-wide methylation profiles reveal quantitative views of human aging rates[J]. **Mol Cell**, 2013, 49(2): 359-367.
- [28] ALFONSO G, GONZALEZ J R. Bayesian neural networks for the optimisation of biological clocks in humans[J]. **BioRxiv**, 2020, 4(21): 052605.
- [29] FREIRE-ARADAS A, GIRON-SANTAMARIA L, MOSQUERA-MIGUEL A, *et al.* A common epigenetic clock from childhood to old age[J]. **Forensic Sci Int Genet**, 2022, 60: 102743.
- [30] AMBROA-CONDE A, GIRON-SANTAMARIA L, MOSQUERA-MIGUEL A, *et al.* Epigenetic age estimation in saliva and in buccal cells[J]. **Forensic Sci Int Genet**, 2022, 61: 102770.
- [31] ALISCH R S, BARWICK B G, CHOPRA P, *et al.* Age-associated DNA methylation in pediatric populations[J]. **Genome Research**, 2012, 22(4): 623-632.
- [32] ROBERTS S, SUDERMAN M, ZAMMIT S, *et al.* Longitudinal investigation of DNA methylation changes preceding adolescent psychotic experiences[J]. **Transl Psychiatry**, 2019, 9(1): 69.