

DOI: 10.13376/j.cbbs/2025043

文章编号: 1004-0374(2025)04-0426-11

DNA甲基化在法医个体年龄推断中的研究进展

张 晶^{1,2#}, 刘 瑞^{1#*}, 米浩源¹

(1 中国人民公安大学侦查学院, 北京 100038; 2 沈阳铁路公安局通辽公安处, 通辽 028000)

摘 要: 未知个体年龄推断是当前法医界的研究热点。DNA 甲基化是目前年龄推断领域最稳定、有效的生物标记物。近年来, 随着检测技术、数据分析方法等迅速发展, 年龄推断的准确性不断提高, 组织适用性不断扩展。本文从 DNA 甲基化年龄相关位点分析及推断模型、转化方法和检测技术、适用的组织类型三个方面进行综述, 旨在梳理 DNA 甲基化的研究成果, 为后续研究提供支持。

关键词: 法医遗传学; 表观遗传学; 年龄推断; DNA 甲基化

中图分类号: Q3-3; R89; DF795.2; D919.2 **文献标志码:** A

Research progress of DNA methylation in forensic individual age prediction

ZHANG Jing^{1,2#}, LIU Rui^{1#*}, MI Hao-Yuan¹

(1 School of Investigation, People's Public Security University of China, Beijing 100038, China; 2 Tongliao Public Security Division of Shenyang Railway Public Security Bureau, Tongliao 028000, China)

Abstract: Inferring the age of an unknown individual is a current research hotspot in forensic science. DNA methylation is currently the most stable and effective biomarker for age prediction. In recent years, with the rapid development of detection technologies, data analysis methods, and other advancements, the accuracy of age prediction has continually improved, and the applicability to various tissues has expanded. This paper provides a review from three aspects: the analysis of DNA methylation age-related CpGs and prediction models, conversion methods and detection technologies, and applicable tissue types to summarize the research findings on DNA methylation and provide support for future studies.

Key words: forensic genetics; epigenetics; age prediction; DNA methylation

未知个体生理年龄推断是法医遗传学表型特征刻画领域的一项重要内容。当犯罪或灾难等现场遗留的生物物证无法与 DNA 数据库中信息匹配时, 若能准确推断其生理年龄, 将对缩小排查范围、提高侦查时效以及各类灾难事件中遇难者的身份鉴定起到重要作用。同时, 个体年龄推断还可以在民事行为能力评估、刑事责任能力评估^[1]、移民、体育竞赛、失踪人口寻找等领域提供有力帮助。

当前, 法医领域进行年龄推断的方法很多, 研究者们也在尝试各类新方法以提高年龄推断准确性和易操作性。法医领域传统的年龄推断, 主要采用身体检查、左手 X 射线检查和牙科检查 (确定牙齿状况和牙列 X 线检查) 相结合的方法, 其被广泛应用于不同人群的生理年龄推断^[2]。但此类方法仅适用

于活体或遗骸存在的情形, 受检材的限制较大^[3], 同时需要法医工作者具有丰富的相关经验。而在实际案件侦破中, 现场可能仅遗留唾液、血液、精斑、头发等生物检材, 此时上述传统方法很难发挥作用。

随着分子生物学和表观遗传学的快速发展, 多种分子生物标记物已被尝试用于生理年龄推断, 包括端粒长度^[4]、DNA 损伤反应^[5]、天冬氨酸外旋

收稿日期: 2024-10-14; 修回日期: 2024-12-28

基金项目: 中国人民公安大学刑事科学技术双一流创新研究专项(2023SYL06)

[#]共同第一作者

*通信作者: E-mail: liurui1987110@126.com

化^[6]、线粒体缺失^[7]和DNA甲基化^[8]等。其中DNA甲基化是个体年龄推断领域目前公认最具应用前景的分子生物标记,因为甲基化DNA的共价键是高度稳定的,已经在被保存超过20年的样本中检测出了甲基化^[9]。DNA甲基化是指在DNA甲基转移酶(DNMT)的催化作用下,S-腺苷-L-甲硫氨酸(SAM)中的甲基转移到DNA胞嘧啶的第5个碳原子上,形成5-甲基胞嘧啶(5-methylcytosine, 5-mC)的过程。由此可知,DNA甲基化并非碱基序列的变化,而是DNA分子上的生物化学修饰。此外,DNA甲基化的检测方法多样,如焦磷酸测序、SNaPshot、甲基化芯片、大规模平行测序和ddPCR等,可在各类法医标准实验室进行检测,以满足各类案件侦破和科研需求。现就DNA甲基化年龄相关位点分析及推断模型、各类检测方法和各类组织中适用情况的研究进展进行综述。

1 DNA甲基化年龄相关位点分析及推断模型

1.1 DNA甲基化年龄相关位点分析

研究者们通过Illumina人类甲基化芯片数据,测定了大量CpG位点与年龄的Pearson(r)或Spearman(ρ)相关性,并筛选出其中相关性较高的位点用于年龄推断。2011年,Bocklandt等^[8]通过使用Illumina 27K人类甲基化芯片对34对男性同卵双胞胎(21~55岁)唾液进行分析,量化了27 578个CpG位点的甲基化程度,确定了88个与年龄显著相关的CpG位点,构建了年龄推断模型,这是首批关于DNA甲基化年龄推断的研究。时隔两年,Horvath^[10]开发了包括353个位点的Horvath表观遗传时钟。同年,Hannum等^[11]基于Illumina 450K甲基化芯片分析了由白种人和西班牙裔人群构成的两个队列(N_1 为482, N_2 为174)血液中的甲基化状态,找到一组对年龄具有强预测能力的甲基化位点,建立的年龄推断模型可实现预测年龄与实际年龄相关性为0.96,平均绝对误差(mean absolute error, MAE)为3.9岁。2024年,Zheng等^[12]基于中国南方250人的衢州队列建立了中国人群DNA甲基化时钟(iCAS-DNAmeAge),并使用中国北方1 070人的CAS(Chinese Aging Score)队列验证此时钟的稳健性,MAE_{Qingzhou}为3.45岁,MAE_{CAS}为4.37岁,这个中国人群的DNA甲基化时钟包含65个CpG位点,其中35个CpG位点的甲基化水平随年龄增加而升高,与蛋白质泛素化有关,其余30个CpG位点的甲基化水平随着衰老而下调,与蛋白质磷酸化有关。另

外,iCAS-DNAmeAge与Horvath时钟、Hannum时钟等共享7个CpG位点,这可能与训练模型的组织类型或人群差异有关^[13]。

1.2 DNA甲基化年龄推断模型

由表观遗传标记DNA甲基化构建的年龄推断模型依赖于数学算法来预测个体的生物年龄。目前甲基化年龄推断领域主要采用多元线性回归模型(multiple linear regression, MLR)、随机森林回归模型(random forest regression, RFR)、支持向量回归模型(support vector machine, SVM)等,国内外学者对各类模型的准确性进行了研究,本文对其进行了综述、比较,具体结果见本段末的表1。邢杨峰等^[14]利用GEO(Gene Expression Omnibus)数据库中的全血样本数据集,分别评估了MLR、RFR、SVM、梯度提升机模型(gradient boosting machine, GBM)和K-最邻近模型(k-nearest neighbors, KNN),结果表明SVM模型效果最好,平均绝对偏差(mean absolute deviation, MAD)值为5.43岁;Ambroa-Conde等^[15]利用SNaPshot技术对DNA甲基化程度进行检测,开发了多元分位数回归模型(multivariate quantile regression, MQR)用于年龄预测,效果较好;Fan等^[16]集成机器学习算法来构建南方汉族的血液表观遗传时钟,用于实际年龄预测,使用不同的机器学习算法建立并评估了四种实际年龄预测模型:逐步回归模型(stepwise regression, SR)、两种SVM模型和RFR模型,测得其MAD值分别为2.97、2.22、2.19和1.29岁,结果表明,无论在女性还是男性数据集中,RFR模型的预测精度最高,MAD为1.29岁;Thong等^[17]使用焦磷酸测序平台,评估了多元变量回归模型(multivariable regression, MVR)和人工神经网络模型(artificial neural network, ANN),证明了ANN模型比MVR具有更高的年龄预测准确性;贾菲等^[18]基于高通量测序平台筛选出血液中与年龄相关性较高的4组位点,评估了ANN、MLR、SVM三种算法,结果表明MLR算法效果最好,MAD为3.685岁;Lau等^[19]利用GEO数据库中的全血样本数据集,分别评估了MLR、RFR、SVM模型、两种ANN模型,结果表明MLR模型优于其余模型,MAD为3.60岁;Levy等^[20]创新了一种用于DNA甲基化分析的自动化模块化深度学习方法MethylNet,此方法可以提取具有生物学意义的特征并实现对个体年龄、细胞类型去卷积、癌症亚型和吸烟状态的较高精度预测。使用该方法对个体生理年龄进行预测,可实现血液中年龄预测

表1 基于不同平台的DNA甲基化年龄推断模型研究

年份	研究者	检测技术	CpG数	样本类型	统计学方法	结果
2023	邢杨峰等 ^[14]	已有数据库(GSE87571 and GSE42861 from Gene Expression Omnibus)	26	血液	多元线性回归模型(MLR)、随机森林回归模型(RFR)、支持向量回归模型(SVM)、梯度提升机模型(GBM)和K-最邻近模型(KNN)	SVM模型效果最好, MAD值为5.43
2022	Ambroa-Conde等 ^[15]	基于SNaPshot技术的富集亚硫酸氢盐测序技术(enrichment bisulfite sequencing)	49	唾液和口腔拭子	多元分位数回归模型(MQR)	MQR模型表现较好, MAD值为3.66
2022	Fan等 ^[16]	焦磷酸测序技术(Pyroseq)	34	血液	逐步回归模型(SR)、支持向量回归模型(SVR-eps和SVR-nu)和随机森林回归模型(RFR)	RFR模型的预测精度最高, MAD值为1.29
2021	Thong等 ^[17]	焦磷酸测序技术(Pyroseq)	29	血液	多变量回归模型(MVR)和人工神经网络模型(ANN)	ANN模型比MVR模型具有更高的年龄预测准确性, MAD值为2.2
2021	贾菲等 ^[18]	靶向下一代测序技术(next generation sequencing, NGS)	26	血液	人工神经网络模型(ANN)、多元线性回归模型(MLR)、支持向量回归模型(SVM)	MLR模型效果最好, MAD值为3.685
2020	Lau等 ^[19]	已有数据库(GSE40279 and GSE42861 from Gene Expression Omnibus)	38	血液	多元线性回归(MLR)、随机森林回归模型(RFR)、支持向量回归(SVM)、具有1个隐藏层(NN1)和2个隐藏层(NN2)的神经网络应用	MLR模型优于大多数复杂的机器学习算法, MAD值为3.60
2020	Levy等 ^[20]	已有数据库(GSE87571)	None	血液	创新了一种用于DNA甲基化分析的自动化模块化深度学习方法(MethylNet)	创新的MethylNet效果较好, MAE值为3.0

MAE 为 3.0 岁的准确性。

2 DNA甲基化的转化方法和检测技术

2.1 DNA甲基化转化方法

目前被认为转化金标准的是亚硫酸氢盐转化法^[21], 即 DNA 中未甲基化的胞嘧啶在亚硫酸氢盐的作用下, 会转化为尿嘧啶, 经 PCR 扩增后, 尿嘧啶转化为胸腺嘧啶, 而此时, 甲基化的胞嘧啶则不会改变, 相当于人为地在 CpG 位点上引入 C/T 变异后进行检测。但亚硫酸氢盐转化法也存在一些不足, 它会导致 DNA 降解、片段化, 可能在应用于微量或降解样本时具有一定局限性, 同时也存在转化不完全的情况, 导致甲基化程度测量结果偏高。此外, 还有酶转化法和亲和富集处理法等。

2.2 DNA甲基化检测技术

从全基因组分析, 到针对特定 CpG 位点的检测, DNA 甲基化的年龄推断精度随着检测技术的发展而不断提升。主要检测技术有: 甲基化检测芯片 (如 Illumina 27K、450K、EPIC 微阵列芯片)、焦磷酸测序、SNaPshot、ddPCR、大规模平行测序和最近新兴的纳米孔测序技术等, 本章节末对其进行了汇总和对比, 见表 2。

甲基化检测芯片技术以美国 Illumina 公司研发的 27K、450K、EPIC 人类甲基化芯片等为代表。基于亚硫酸氢盐转化法, Illumina 公司设计了 Infinium I 和 II 两种探针, Infinium I 探针基于单碱基延伸原理, 使用 U 型磁珠 (尾部为 A, 用来检测非甲基化位点 T) 和 M 型磁珠 (尾部为 G, 用来检测甲基化位点 C), 通过两种不同探针的信号比值来测定甲基化程度, 而 Infinium II 探针基于普通延伸反应, 只使用一种磁珠 (末端为 C, 与 CpG 位点的 G 碱基结合), 仅向前延伸一个碱基, 通过延伸的 A 或 G 碱基 (分别对应非甲基化和甲基化位点) 的信号比值计算目的位点的甲基化程度。其中 27K 芯片使用 Infinium I 型探针, 而 450K 和 EPIC 芯片中同时采用 I 和 II 两种探针, 使得检测位点范围进一步扩大。甲基化芯片技术价格较为昂贵, 但具有位点覆盖率高、高通量等优点, 适用于年龄相关性较高的位点筛选以及表型特征刻画等其他领域。Bocklandt 等^[8]通过 Illumina 27K 甲基化芯片数据进行了基于唾液的生理年龄推断, Hannum 等^[11]使用 450K 甲基化芯片数据在白种人和西班牙裔中构建了血液中的年龄推断模型, MAE 为 3.9 岁。Lau 等^[19]使用基于 450K 平台的公开数据集, 筛选了血液中与年龄显

著相关的 16 个 CpG 位点, 并训练 MLR 模型, MAD 为 3.76 岁。

焦磷酸测序法是基于酶联发光反应测定焦磷酸盐 (PPi) 的实时 DNA 测序技术, 能够做到“边合成边测序”并实现定量检测。该法不仅测序时间较短, 还能评估亚硫酸氢盐转化是否完全, 预防假阳性数据, 保证结果的可靠性, 被认为是甲基化检测的“金标准”。Eipel 等^[22]基于焦磷酸测序技术, 仅使用 3 个 CpG 位点就实现了口腔拭子的年龄推断, MAD 为 4.3 岁。但是, 焦磷酸测序仅适用于单位点分析, 无法同时分析不同位点的甲基化程度^[23]。

SNaPshot 技术基于单碱基延伸原理, 由 DNA 聚合酶、四种荧光标记的 ddNTP、5' 端紧靠 CpG 位点的引物和亚硫酸氢盐转化后的 DNA 模板构成 PCR 反应体系, 引物延伸一个碱基即终止反应, DNA 的甲基化水平可根据不同峰之间的荧光比值来确定^[24]。此技术可以进行多重甲基化位点分析, 且灵敏度较高, 但不同荧光染料间荧光强度不平衡可能影响测量的准确性, 且检测周期较长 (2~3 d)^[24]。

EpiTYPER 技术, 又称飞行质谱法 (MassArray), 经亚硫酸氢盐转化后的 DNA 模板, 经过 PCR 扩增后, 可通过飞行时间质谱检测产物序列上的 A/G 差异, 两种 CpG 位点的分子量相差 16 Da。该技术可覆盖长 400~600 bp 的区域位点, 准确性高, 数据简便。李姍飞等^[25]基于 EpiTYPER 技术筛选出 8 个与年龄高度相关的 CpG 位点, 实现了中国北方汉族男性血液 DNA 的年龄推断, MAD 为 2.69 岁。该方法虽然灵敏度和准确度高, 但实验操作较为烦琐, 且成本较高。

ddPCR (微滴数字 PCR) 技术是近年来新兴的单分子绝对定量检测技术, 该技术将待检测的亚硫酸氢盐转化后的 DNA 模板分子, 分割为数万个单独的反应体系, 理论上可实现“单分子扩增”。ddPCR 技术是目前唯一可用于多重检测且绝对定量的 DNA 甲基化检测技术。Han 等^[26]的研究表明, ddPCR 检测结果甚至略优于焦磷酸测序结果, 可以进一步提高 DNA 甲基化程度测量的准确性和灵敏度。Zhou 等^[27]基于 ddPCR 平台构建了一种可同时复合检测 4 个 CpG 位点的甲基化检测体系, 使用随机森林法建立血液中的年龄推断模型, MAE 为 4.5923 岁, 该体系在亚硫酸氢盐转化后 DNA 投入量为 10~50 ng 时效果最佳。

大规模平行测序 (MPS), 也称下一代测序技术 (NGS), 可以实现同时对数百万个 DNA 片段的平

表2 DNA甲基化检测方法的比较

类别	技术	优点	局限性
绝对定量DNA 甲基化测定	大规模平行测序	边测序边分析, 灵敏度高, 可检测未知序列	检测片段较短, 技术成本较高
	ddPCR(微滴数字PCR)	可用于多重检测, 准确性和灵敏度高	可同时检测位点数较少, 通量有待提高
	基于SNaPshot技术的富集 亚硫酸氢盐测序技术	可并行富集, 灵敏度高	准确性待验证, 技术成本高, 检测周期长
相对定量DNA 甲基化测定	焦磷酸测序技术	灵敏度高、重复性好、操作简便, 可精确检测单个CpG位点甲基化比例	无法同时分析不同位点的甲基化程度
	全基因组重亚硫酸盐测序	可检测基因组中全部的甲基化位点	成本高, 步骤烦琐, 特定位点的检测准确性不高
	EpiTYPER技术 甲基化芯片技术	可检测片段长度长, 准确性高, 灵敏度高 可同时检测人类基因组中数十万个甲基化位点, 覆盖范围广, 操作简便	性价比比较低, 超出检测质量范围的位点无法测定 成本高, 只能检测已知位点

行序列测定。贾菲等^[18]通过 Miseq FGX 测序平台 (MPS) 基于 26 个位点建立了适用于血液、血液斑和唾液斑的年龄推断模型, MAD 值分别为 3.685 岁、1.25 岁、1.79 岁, 灵敏度为 100 ng。王紫薇等^[28]使用 NGS 技术和焦磷酸测序, 对 Pan 等^[29]基于 SNaPshot 技术开发的年龄推断模型在华东汉族人群中跨平台应用进行研究, 发现使用 MPS 技术建立的年龄推断模型准确度 (MAD 为 4.41 岁) 高于焦磷酸测序平台 (MAD 为 4.81 岁), 证实了 MPS 技术在甲基化分析和法医学中的实用性。MPS 技术可同时分析上千个甲基化位点, 具有速度快、准确性高等优势, 但这种方法价格较为昂贵, 且数据分析较为复杂^[24]。

3 DNA甲基化年龄推断在各类组织中的应用

DNA 甲基化具有组织特异性, 即同一个位点的甲基化模式在不同组织间会产生差异。因此, 基于某一组织开发的年龄推断模型在应用于其他组织时, 可能会一定程度上影响其年龄推断的准确性; 同时也存在一些位点在唾液、血液、骨组织中的甲基化模式相似, 可用于多组织年龄推断。因此, 这里列举了近年来在单一组织和多种组织中甲基化年龄推断模型的研究情况, 望为研究者们提供参考。

3.1 血液

血液及其形成的血痕, 是犯罪现场最容易发现、最为常见、支撑破案量最大的生物物证, 血液也被认为是蕴藏着 DNA 信息的最权威检材, 因而国内外学者对基于血液样本的 DNA 甲基化年龄推断进行了大量研究, 本文对研究成果进行了总结, 具体结果见本段末的表 3。Feng 等^[30]基于 EpiTYPER 平台建立了中国汉族男性血液中适用的 9-CpG 年龄推断模型, MAD 为 2.89 岁; 孙晓萌等^[31]对上述 9-CpG 年龄推断模型在不同海拔地域和不同性别人群中年龄推断的适用性进行验证及优化, 采集 367 份血液样本, 使用 EpiTYPER 技术检测 9 个 CpG 位点的甲基化值, 加入平原和高原变量后, 新建立 8-CpG 模型, MAD 为 2.90 岁; 韩雪丽^[32]基于 SNaPshot 技术对 374 例血液样本的 DNA 甲基化水平进行了检测, 对 8 个基因的 CpG 位点进行了检测, 并构建 SVR 年龄推断模型, MAE 为 2.88 岁; 高林林等^[33]基于 EpiTYPER 平台对前期已构建的基于 DNA 甲基化的年龄推断模型进行优化, 使用 196 份汉族人群血液样本, MAD 值为 2.95 岁; 王紫薇等^[28]使用焦磷酸测序技术和靶向下一代测序技术分别对 48

表3 基于血液样本的DNA甲基化年龄推断研究进展

年份	研究者	技术	样本数	基因	CpG位点	准确性(岁)
2018	Feng等 ^[30]	EpiTYPER	390	TRIM59	chr3:160450189/92/99	2.9
				TRIM59	chr3:160450299	
				RASSF5	chr1:206512090	
				C1orf132	chr1:207823715	
				--	chr10:22334463/65	
				PDE4C	chr19:18233105	
				PDE4C	chr19:18233127/31/33	
				CCDC102B	chr18:68722210	
				ELOVL2	chr6:11044628/31/34	
2021	孙晓萌等 ^[31]	EpiTYPER	367	TRIM59	chr3:160450189/92/99	4.0
				TRIM59	chr3:160450299	
				RASSF5	chr1:206512090	
				C1orf132	chr1:207823715	
				--	chr10:22334463/65	
				PDE4C	chr19:18233105	
				PDE4C	chr19:18233127/31/33	
				CCDC102B	chr18:68722210	
				ELOVL2	chr6:11044628/31/34	
2022	韩雪丽 ^[32]	SNaPshot	374	ELOVL2	chr6:11044628	3.7
				FHL2	cg06639320	
				C1orf132	chr1:207823723	
				CCDC102B	cg19283806	
				KLF14	cg14361627	
				SYNE2	cg17740900	
				TRIM59	cg07553761	
				--	cg26947034	
2022	高林林等 ^[33]	EpiTYPER	196	同Feng等 ^[30]	同Feng等 ^[30]	2.9
2023	王紫薇等 ^[28]	Pyroseq NGS	48	ITGA2B	chr17:44390358	4.8
				KLF14	cg14361627	4.4
				FHL2	cg06639320	
				ZNF423	cg04208403	
				ASPA	cg02228185	
				CCDC102B	cg19283806	
2023	Yang等 ^[34]	Pyroseq	241	C1orf132	cg10501210	<3
				ELOVL2	cg16867657	
				FHL2	cg06639320	
				KLF14	cg14361627	
				NPTX2	cg00548268	
				TRIM59	chr3:160450199	
2023	Varshavsky等 ^[35]	WGBS	11 910	ELOVL2	cg16867657	2.1
				FHL2	cg22454769	
				FHL2	cg06639320	
				OTUD7A	cg04875128	
				CCDC102B	cg19283806	
				ELOVL2	cg24724428	
				TRIM59	cg07553761	
				FHL2	cg24079702	
				RASSF5	cg08128734	
				GRM2	cg12934382	

表3 基于血液样本的DNA甲基化年龄推断研究进展(续表)

年份	研究者	技术	样本数	基因	CpG位点	准确性(岁)
				--	cg08468401	
				CC2D2A	cg20816447	
				ZEB2	cg00573770	
				ZYG11A	cg06335143	
				PMPCB	cg06155229	
				--	cg03032497	
				PDZK1IP1	cg06619077	
				TP73	cg17804348	
				IGSF11	cg00329615	
				March11	cg23479922	
				--	cg10501210	
				TIAL1	cg19991948	
				SORBS1	cg27312979	
				CD44	cg23186333	
				--	cg25413977	
				FAM171A2	cg22078805	
				RNF180	cg17621438	
				ADAMTS6	cg21878650	
				ANKRD11	cg04503319	
				EDARADD	cg09809672	

--代表基因未命名。

例血液样本的 DNA 甲基化水平进行了检测，筛选出 6 个 DNA 甲基化程度随年龄变化的 CpG 位点(位于 *ITGA2B*、*KLF14*、*FHL2*、*ZNF423*、*ASPA*、*CCDC102B* 基因),MAD 值分别为 4.81 岁和 4.41 岁; Yang 等^[34] 基于焦磷酸测序技术和 RFR 建立了年龄预测模型,检测了 241 份血迹样本的 DNA 甲基化水平, MAD 值小于 3 岁; Varshavsky 等^[35] 基于全基因组重亚硫酸盐测序技术对 11 910 个血液样本 DNA 甲基化组进行机器学习分析,回归模型在预测年龄方面的中位误差为 2.1 岁。

综上,多种检测方法及数据处理方法在基于血液样本的 DNA 甲基化年龄推断上进行了大量应用,使得年龄预测模型更准确,未来将针对特定 CpG 位点的 DNA 甲基化程度测定及模型构建进行更广泛的研究。

3.2 唾液和口腔拭子

除血液外,唾液也是犯罪现场较常见的生物物证,尤其是在有关性侵的犯罪现场,唾液样本极易提取,本文对国内外学者研究成果进行了总结,具体结果见本段末的表 4。林培双^[36] 基于国产微滴数字 PCR (droplet digital PCR, ddPCR) 平台进行 147 份唾液样本中年龄相关 CpG 位点检测,建立了 DNA 甲基化年龄推断模型, MAD 值为 3.43 岁; Poussard

等^[37] 收集并分析了 115 名年龄在 0~88 岁之间的法国人的唾液和口腔拭子样本,基于优化后的焦磷酸测序技术建立了年龄预测模型,唾液、口腔拭子的 MAD 值分别为 3.5 岁和 3.9 岁; Schwender 等^[38] 基于焦磷酸测序技术结合逐步回归模型,对 141 个口腔拭子样本的 DNA 甲基化程度进行了检测, MAD 为 5.33 岁,证明了口腔拭子作为血液样本的合适替代品的合理性。

3.3 精液

精液中的 DNA 甲基化模式与血液、唾液不同,但在涉及性犯罪的案件中,精液、精斑样本存在较为普遍,本文对国内外学者研究成果进行了总结,具体结果见本段末的表 5。Xiao 等^[39] 基于 SNaPshot 技术分析了精液中 22 个年龄相关的 CpG 位点,其中 cg21843517 位点表现出最强的年龄相关性(r 为 0.853),并训练了精液中年龄推断模型,效果较好; Li 等^[40] 基于焦磷酸测序技术结合 MLR 模型对 38 份精液样本进行了检测,建立了年龄推断模型,能够准确预测汉族男性精液捐献者的年龄; Heidegger 等^[41] 使用大规模平行测序技术对两个精液样本的 DNA 甲基化程度进行了检测,预测年龄和实际年龄之间的差异在所建立模型的预期范围内; Pisarek 等^[42] 基于大规模平行测序技术,建立了一种年龄

表4 基于唾液样本的DNA甲基化年龄推断研究进展

年份	研究者	技术	样本数	基因	CpG位点	准确性(岁)
2022	林培双 ^[36]	ddPCR	147	PTPN7	cg18384097	3.4
				SST	cg00481951	
				CNGA3	cg19671120	
				KLF14	cg14361627	
				TSSK6	cg08928145	
				TBR1	cg12757011	
				SLC12A5	cg07547549	
				ELOVL2	chr6: 11044628	
				FHL2	chr2: 105399282	
				MIR29B2C	chr1: 207823681	
				TRIM59	chr3: 160450189	
2023	Poussard等 ^[37]	Pyroseq	115	ELOVL2	chr6: 11044628	3.5
				FHL2	chr2: 105399282	
				KLF14	chr7: 130734355	
				MIR29B2C	chr1: 207823681	
				TRIM59	chr3: 160450189	
2021	Schwender等 ^[38]	Pyroseq	141	PDE4C	cg17861230	5.3
				ELOVL2	cg16867657	
				ITGA2B	cg25809905	
				ASPA	cg02228185	
				EDARADD	cg09809672	
				SST	cg00481951	
				KLF14	cg08097417	
				SLC12A5	cg07547549	

推断模型, 预测年龄的平均绝对误差为 5.1 年, 研究揭示进一步提高精液 DNA 年龄预测的准确性需要技术进步。

3.4 多组织

血液、唾液等组织中一些位点的甲基化模式已有研究证明是相似的。贾菲等^[18]在血液中建立的年龄推断模型 (MAD 为 3.685 岁), 在唾液斑迹中进行验证, MAD 为 1.79 岁。Li 等^[43]在血液中建立的 6-CpG 年龄推断模型, MAD 为 2.72 岁, 还将这 6 个位点在唾液样本中进行训练, 得到的年龄推断模型 MAD 为 2.1 岁。

Lee 等^[44]收集了 20 名去世韩国人的 180 份死后组织样本 (18~78 岁), 使用 Illumina EPIC 甲基化芯片检测其甲基化程度, 使用 6 种表观遗传时钟 (Pan-tissue clock^[10]、Hannum clock^[11]、Skin and blood clock^[45]、Zhang clock^[46] (Elastic Net & BLUP)、MEAT clock^[47, 48]、PhenoAge^[49]) 计算了各种组织样本中的预测年龄, 这是第一个将多种表观遗传时钟应用于同一个体的各种组织类型的研究, 对表观遗传时钟在死后不同组织类型的适用性提供了有价值

的见解, 展示了它们在法医应用中的潜在可能性, 虽然这些时钟最初是基于欧洲样本开发的, 但此研究证实了它们也适用于韩国个体死后提取的各种组织类型。

3.5 其他组织类型

法医现场的物证类型多样, 研究者们还对头发毛囊、指甲等组织的年龄推断进行了研究。Hao 等^[50]基于多重甲基化 SNaPshot 方法, 使用 10 个 CpG 位点在头发毛囊中构建了 MLR 年龄推断模型, MAD 为 3.68 岁。Fokias 等^[51]建立了适用于指甲和脚趾甲的年龄推断模型, MAD 为 5.48 岁, 指 (趾) 甲具有较强的抗腐蚀性, 且易于采样, 在年龄推断领域具有较强的适用性。Voisin 等^[47]基于 Illumina 27K、450K、EPIC 甲基化芯片的 12 个独立数据集的 682 个骨骼肌肉样本的 DNA 甲基化数据, 开发了包含 200 个 CpG 位点的骨骼肌组织中适用的表观遗传时钟 (Muscle Epigenetic Age Test clock), MAE 为 4.6 岁, 该表观遗传时钟可在 GitHub 上免费使用, 其中 16 个 CpG 位点与 Pan-tissue clock 中的 353 个 CpG 位点相同。Jung 等^[52]基于 MPS 平台, 在 85 例已

表5 基于精液样本的DNA甲基化年龄推断研究进展

年份	研究者	技术	样本数	基因	CpG位点	MAD
2023	Xiao等 ^[39]	SNaPshot	90	None	cg19998819	2.2
				SNRNP35	cg21843517	
				LINC00423	cg18037145	
				COL18AI	cg04123357	
				LINC00703	cg12277678	
				None	cg25187042	
				C12orf4	cg11262154	
				None	cg03634854	
				GIT1	cg13872326	
				BHLHA9	cg20828122	
				PTPRTPT	cg20602007	
				None	cg27231587	
				PPICA	cg01789162	
				PURA	cg27111970	
				DLG2	cg04119405	
				None	cg24812634	
				COMP	cg19983027	
				LYPLA1	cg03030301	
				SLAMF6	cg25715498	
				NOX4c	cg06979108	
				TTC7B	cg06304190	
				None	cg12837463	
2020	Li等 ^[40]	Pyroseq	38	NOX4	cg06979108	4.1
					cg12837463	
2022	Heidegger等 ^[41]	NGS	2	NOX4	cg06979108	5.1
2021	Pisarek等 ^[42]	NGS	40	PPP2R2C	cg02766173	5.1
				EXOC3	cg10528482	
				SH2B2	cg00018181	
				IFITM2	cg01886988	
				SYT7	cg17147820	
				ARHGEF17	cg09855959	
				TUBB3	cg18701351	
				TBX4	cg19862839	
				GALR2	cg07909178	
				PALM	cg17704154	

故韩国人肋骨 DNA 中测定了 45 个 CpG 位点的甲基化状态，并用其中 9 个 CpG 位点进行 MLR 年龄推断模型构建，MAE 为 4.97 岁，此模型在欧洲 181 个肋软骨样本中也表现出较好的年龄推断性能，MAE 为 6.06 岁，灵敏度可达 5.5 ng，此模型在大规模灾害等人类遗骸鉴定方面具有较大应用潜力。

4 结语

随着 DNA 甲基化检测技术的发展，年龄推断的精确度不断提升，适用的组织类型更加全面，检测时效性也不断提高。未来需要探索效果更好的甲

基化转化试剂，以降低 DNA 在转化过程中的损耗和总投入量，提高灵敏度，这在实际案件侦破中具有重要意义。研究重心应向开发多组织或跨平台通用的年龄推断模型发展，减小或消除因不同组织、不同检测平台间的差异而导致的年龄推断误差，以期更好地服务法医实践，为刑事侦查提供更加精确的线索指引。

[参 考 文 献]

[1] Schmeling A, Olze A, Reisinger W, et al. Age estimation of living people undergoing criminal proceedings. Lancet,

- 2001, 358: 89-90
- [2] Schmelting A, Geserick G, Reisinger W, et al. Age estimation. *Forensic Sci Int*, 2007, 165: 178-81
- [3] 敖健恒, 董文娟, 吕孟然, 等. DNA甲基化的功能特点及法医学应用的研究进展. *中国优生与遗传杂志*, 2023, 31: 2183-8
- [4] Srettabunjong S, Satitsri S, Thongnoppakhun W, et al. The study on telomere length for age estimation in a Thai population. *Am J Forensic Med Pathol*, 2014, 35: 148-53
- [5] Slijepcevic P. DNA damage response, telomere maintenance and ageing in light of the integrative model. *Mech Ageing Dev*, 2008, 129: 11-6
- [6] Rajkumari S, Nirmal M, Sunil P M, et al. Estimation of age using aspartic acid race misation in human dentin in Indian population. *Forensic Sci Int*, 2013, 228: 38-41
- [7] Cortopassi GA, Shibata D. A pattern of accumulation of a somatic deletion of mitochondrial DNA in aging human tissues. *Proc Natl Acad Sci USA*, 1992, 89: 7370-4
- [8] Bocklandt S, Lin W, Sehl ME, et al. Epigenetic predictor of age. *PLoS One*, 2011, 6: e14821
- [9] Silva DSBS, Antunes J, Balamurugan K, et al. Developmental validation studies of epigenetic DNA methylation markers for the detection of blood, semen and saliva samples. *Forensic Sci Int Genet*, 2016, 23: 55-63
- [10] Horvath S. DNA methylation age of human tissues and cell types. *Genome Biol*, 2013, 14: R115
- [11] Hannum G, Guinney J, Zhao L, et al. Genome-wide methylation profiles reveal quantitative views of human aging rates. *Mol Cell*, 2013, 49: 359-7
- [12] Zheng Z, Li J, Liu T, et al. DNA methylation clocks for estimating biological age in Chinese cohorts. *Protein Cell*, 2024, 15: 575-93
- [13] Bell CG, Lowe R, Adams PD, et al. DNA methylation aging clocks: challenges and recommendations. *Genome Biol*, 2019, 20: 249
- [14] 邢杨峰, 冀志敏, 李俊丽, 等. 机器学习技术结合Y染色体CpG位点推断男性年龄. *中国法医学杂志*, 2023, 38: 381-4+389
- [15] Ambroa-Conde A, Girón-Santamaría L, Mosquera-Miguel A, et al. Epigenetic age estimation in saliva and in buccal cells. *Forensic Sci Int Genet*, 2022, 61: 102770
- [16] Fan H, Xie Q, Zhang Z, et al. Chronological age prediction: developmental evaluation of DNA methylation-based machine learning models. *Front Bioeng Biotechnol*, 2022, 9: 8199991
- [17] Thong Z, Tan JYY, Loo ES, et al. Artificial neural network, predictor variables and sensitivity threshold for DNA methylation-based age prediction using blood samples. *Sci Rep*, 2021, 11: 1744
- [18] 贾菲, 孙学科, 喻少波, 等. DNA甲基化水平的法医学个体年龄推断. *中国法医学杂志*, 2021, 36: 373-8
- [19] Lau PY, Fung WK. Evaluation of marker selection methods and statistical models for chronological age prediction based on DNA methylation. *Leg Med (Tokyo)*, 2020, 47: 101744
- [20] Levy JJ, Titus AJ, Petersen CL, et al. MethylNet: an automated and modular deep learning approach for DNA methylation analysis. *BMC Bioinformatics*, 2020, 21: 108
- [21] Cao Y, Bai Y, Yuan T, et al. Single-cell bisulfite-free 5mC and 5hmC sequencing with high sensitivity and scalability. *Proc Natl Acad Sci USA*, 2023, 120: e2310367120
- [22] Eipel M, Mayer F, Arent T, et al. Epigenetic age predictions based on buccal swabs are more precise in combination with cell type-specific DNA methylation signatures. *Aging*, 2016, 8: 1034-48
- [23] Huang Y, Qu S, Xiao Y, et al. Progress in age estimation based on DNA methylation. *J Forensic Sci Med*, 2023, 9: 360-6
- [24] Kaminsky Z, Petronis A. Methylation SNaPshot: a method for the quantification of site-specific DNA methylation levels. *Methods Mol Biol*, 2009, 507: 241-55
- [25] 李姗姗, 王玲, 赵慧, 等. DNA甲基化检测与生物物证时间关联性研究进展. *刑事技术*, 2019, 44: 337-42
- [26] Han Y, Franzen J, Stiehl T, et al. New targeted approaches for epigenetic age predictions. *BMC Biol*, 2020, 18: 71
- [27] Zhou Y, Wang Y, Song M, et al. A high-throughput droplet digital PCR system aiming eight DNA methylation targets for age prediction. *J Pharm Biomed Anal*, 2024, 240: 115943
- [28] 王紫薇, 李成涛, 刘希玲. DNA甲基化年龄推断模型在华东汉族人群中的跨平台应用. *法医学杂志*, 2023, 39: 441-6
- [29] Pan C, Yi S, Xiao C, et al. The evaluation of seven age-related CpGs for forensic purpose in blood from Chinese Han population. *Forensic Sci Int Genet*, 2020, 46: 102251
- [30] Feng L, Peng F, Li S, et al. Systematic feature selection improves accuracy of methylation-based forensic age estimation in Han Chinese males. *Forensic Sci Int Genet*, 2018, 35: 38-45
- [31] 孙晓萌, 刘雅洁, 李姗姗, 等. 9-CpG年龄推断在不同海拔地域和性别人群中的验证及优化. *中国法医学杂志*, 2021, 36: 274-9
- [32] 韩雪丽. 基于8个强年龄相关性CpGs的汉族人群血样个体年龄推断研究[D]. 武汉: 华中科技大学, 2023
- [33] 高林林, 王科文, 李佑英, 等. 基于血斑的9-CpG年龄推断在浙江汉族中的验证及优化. *刑事技术*, 2022, 47: 627-32
- [34] Yang F, Qian J, Qu H, et al. DNA methylation-based age prediction with bloodstains using pyrosequencing and random forest regression. *Electrophoresis*, 2023, 44: 835-44
- [35] Varshavsky M, Harari G, Glaser B, et al. Accurate age prediction from blood using of small set of DNA methylation sites and a cohort-based machine learning algorithm. *bioRxiv*, 2023: 2023.01.20.524874.
- [36] 林培双. 基于国产微滴数字PCR的法医DNA应用研究[D]. 贵阳: 贵州医科大学, 2023
- [37] Poussard A, Curci JY, Siatka C, et al. Evaluation of DNA methylation-based age-prediction models from saliva and buccal swab samples using pyrosequencing data. *Forensic Sci*, 2023, 3: 192-204
- [38] Schwender K, Holländer O, Klopffleisch S, et al. Development of two age estimation models for buccal swab samples based on 3 CpG sites analyzed with

- pyrosequencing and minisequencing. *Forensic Sci Int Genet*, 2021, 53: 102521
- [39] Xiao C, Li Y, Chen M, et al. Improved age estimation from semen using sperm-specific age-related CpG markers. *Forensic Sci Int Genet*, 2023, 67: 102941
- [40] Li L, Song F, Lang M, et al. Methylation-based age prediction using pyrosequencing platform from seminal stains in Han Chinese males. *J Forensic Sci*, 2020, 65: 610-9
- [41] Heidegger A, Pisarek A, de la Puente M, et al. Development and inter-laboratory validation of the VISAGE enhanced tool for age estimation from semen using quantitative DNA methylation analysis. *Forensic Sci Int Genet*, 2022, 56: 102596
- [42] Pisarek A, Pośpiech E, Heidegger A, et al. Epigenetic age prediction in semen – marker selection and model development. *Aging (Albany NY)*, 2021, 13: 19145-64
- [43] Li X, Li W, Xu Y. Human age prediction based on DNA methylation using a gradient boosting regressor. *Genes*, 2018, 9: 424
- [44] Lee JM, Park SU, Lee SD, et al. Application of array-based age prediction models to post-mortem tissue samples. *Forensic Sci Int Genet*, 2024, 68: 102940
- [45] Horvath S, Oshima J, Martin GM, et al. Epigenetic clock for skin and blood cells applied to Hutchinson Gilford Progeria Syndrome and *ex vivo* studies. *Aging*, 2018, 10: 1758-75
- [46] Zhang Q, Vallergera CL, Walker RM, et al. Improved precision of epigenetic clock estimates across tissues and its implication for biological ageing. *Genome Med*, 2019, 11: 54
- [47] Voisin S, Harvey NR, Haupt LM, et al. An epigenetic clock for human skeletal muscle. *J Cachexia Sarcopenia Muscle*, 2020, 11: 887-98
- [48] Voisin S, Jacques M, Landen S, et al. Meta-analysis of genome-wide DNA methylation and integrative omics of age in human skeletal muscle. *J Cachexia Sarcopenia Muscle*, 2021, 12: 1064-78
- [49] Levine ME, Lu AT, Quach A, et al. An epigenetic biomarker of aging for lifespan and healthspan. *Aging*, 2018, 10: 573-91
- [50] Hao T, Guo J, Liu J, et al. Predicting human age by detecting DNA methylation status in hair. *Electrophoresis*, 2021, 42: 1255-61
- [51] Fokias K, Dierckx L, Van De Voorde W, et al. Age determination through DNA methylation patterns in fingernails and toenails. *Forensic Sci Int Genet*, 2023, 64: 102846
- [52] Jung JY, So MH, Jeong K, et al. Epigenetic age prediction using costal cartilage for the investigation of disaster victims and missing persons. *J Forensic Sci*, 2024, 69: 1578-86