

# Compact DNA methylation panels for forensic age estimation: cross-cohort evaluation and calibration using public whole-blood data

**Baihereye Abudouwaier**

Shanxi Medical University

**Shengjie Gao**

Shenzhen University

**Halimureti Simayijiang**

**halimureti@qq.com**

Shenzhen University

---

## Research Article

**Keywords:** DNA methylation, forensic age estimation, chronological age, forensic epigenetics, whole blood, model calibration

**Posted Date:** August 27th, 2026

**DOI:** <https://doi.org/10.21203/rs.3.rs-10659770/v1>

**License:**   This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

**Additional Declarations:** No competing interests reported.

---

# Abstract

DNA methylation-based age estimation is promising for forensic intelligence, but transferability across cohorts and platforms remains a major challenge. We evaluated whether compact, literature-supported blood CpG panels retain predictive performance across independent public cohorts and whether limited marker expansion or local calibration improves accuracy. A health-filtered GSE207605-derived resource was used for model development after excluding GSE40279 and GSE69270, leaving 8,986 samples for nested group-aware cross-validation. Three Elastic Net models containing four, eight, or nine CpGs were compared and externally evaluated in GSE40279 ( $n = 656$ ), GSE69270 ( $n = 184$ ), and GSE141682 ( $n = 42$ ). Internal mean absolute error was 5.97, 5.86, and 5.84 years for the four-, eight-, and nine-CpG models, respectively. Raw external performance was cohort dependent: the four-CpG model achieved the lowest MAE in GSE40279 (8.62 years) and GSE69270 (3.90 years), whereas the eight-CpG model performed marginally better in GSE141682 (7.85 years). In leakage-free repeated-split calibration, intercept-only adjustment reduced the eight-CpG model mean held-out MAE from 8.83 to 5.56 years, 4.28 to 2.46 years, and 7.77 to 3.92 years, respectively. These findings support compact blood-based panels as the primary forensic strategy and indicate that cohort-specific calibration may be more consequential than adding markers. Prospective laboratory validation remains required before casework implementation.

## 1. Introduction

DNA methylation-based age estimation has become an important tool in forensic epigenetics because age-associated CpG methylation can provide investigative information when the identity of a biological trace donor is unknown [4–6, 13–19]. For operational use, however, predictive accuracy is only one requirement. Forensic assays must also remain compatible with limited or degraded DNA, targeted laboratory workflows, manageable marker numbers, transparent interpretation, and reproducible performance across laboratories and populations [4–6, 13–16]. These constraints explain the continuing interest in compact blood-based marker sets centered on loci such as ELOVL2, FHL2, PDE4C, EDARADD, KLF14, ASPA, MIR29B2CHG/C1orf132, and TRIM59.

Although many of these loci show strong age associations, published models frequently differ in marker composition, platform, population, preprocessing, and calibration strategy. A model may therefore preserve a high correlation with chronological age while showing substantial absolute error or systematic bias in an external cohort. This distinction is particularly important for forensic implementation because casework requires not only ranking individuals by age but producing age estimates whose error and bias are understood under the laboratory and population conditions in which the assay will be used [20–23].

Large public whole-blood methylation resources provide an opportunity to study this transferability problem while maintaining a common tissue context [2, 9–12, 24]. Such resources also contain age-associated CpGs discussed in the broader epigenetic-aging literature [1, 3, 7, 8]. In the present study, that broader literature is used only as secondary context for controlled marker expansion; the primary endpoint throughout is forensic chronological-age estimation. Restricting the analysis to whole blood reduces tissue-related heterogeneity, although array platform, cohort composition, blood-cell composition, and preprocessing history remain potential sources of between-cohort variation [2, 3].

Accordingly, we evaluated whether a compact literature-supported forensic CpG core remains competitive when modestly expanded within a nine-CpG candidate set, and whether local calibration can reduce cohort-specific offset without information leakage. A health-filtered GSE207605-derived resource was used for nested group-aware model development after removal of cohorts reserved for external testing. Four-, eight-, and nine-CpG Elastic Net models

were compared in three independent external whole-blood cohorts, including a healthy Han Chinese forensic-relevant dataset. We hypothesized that limited marker expansion would yield only modest gains unless transferability and calibration were explicitly addressed.

## **2. Forensic marker framework**

### **2.1. Compact forensic core and controlled expansion**

Blood-based forensic age-estimation studies repeatedly identify a relatively small group of CpG loci with practical value, including ELOVL2, FHL2, PDE4C, KLF14, EDARADD, ASPA, MIR29B2CHG/C1orf132, and TRIM59 [4–6, 13, 15, 17, 18]. We therefore treated compactness and literature support as primary design considerations rather than optimizing marker number without constraint. Model A represented a prespecified four-CpG forensic core. Models B and C evaluated controlled expansion within the supplied nine-CpG feature set so that any gain in predictive performance could be weighed against additional assay complexity.

### **2.2. Operational marker zones**

For descriptive purposes, candidate markers were organized into three operational zones. Zone I comprised loci prioritized for compact chronological-age prediction; Zone II comprised additional age-associated loci considered for controlled expansion and transferability testing; and Zone III represented broader biological-deviation signals that were not treated as a primary forensic endpoint. This zoning was used only as an organizational framework for panel development and should not be interpreted as a validated biological classification. The empirical models in this study were evaluated only for chronological-age prediction.

## **3. Materials and Methods**

### **3.1. Study design**

This study was a forensic-oriented secondary analysis of public whole-blood DNA methylation data with a targeted panel-development objective (Fig. 1). The workflow combined (i) curation of repeatedly reported forensic age-associated loci; (ii) selection and harmonization of public whole-blood methylation datasets; (iii) evaluation of a supplied nine-CpG candidate feature set; (iv) comparison of a prespecified compact four-CpG core with controlled eight- and nine-CpG expansions; and (v) cohort-aware internal and external validation with a prespecified calibration sensitivity analysis. The primary outcome was chronological-age prediction; broader biological-age interpretation was treated as secondary context only.

### **3.2. Public dataset selection**

To minimize tissue-related confounding, analyses were restricted to public whole-blood DNA methylation datasets. This blood-based restriction was adopted as a practical analytical choice to improve comparability, rather than to suggest that blood is the only relevant sample type for forensic or translational age assessment. Datasets were considered eligible if they met the following criteria: (i) blood origin; (ii) public availability through NCBI GEO; (iii) availability of chronological age information at the sample or series level; (iv) sufficient sample size for discovery, comparison, or external evaluation; and (v) compatibility with array-based methylation analysis. Based on these criteria, the primary datasets selected for the present framework were GSE207605, GSE40279, GSE141682, and GSE69270 [2, 9–12]. Cohort-level dataset annotations and inclusion/exclusion decisions are provided in Online

Resource 1, Table S1, and sample-level cohort, age, sex/gender coding, and model predictions are provided in Online Resource 2.

### 3.3. Public datasets used in the framework

GSE207605 was used as the primary harmonized training resource. GEO describes it as a third-party reanalysis of 11,910 whole-blood methylomes with processed beta values and donor ages. The supplied health-filtered model-feature table contained 9,831 records across nine candidate CpGs; five incomplete records were removed, leaving 9,826 complete cases. To prevent overlap with held-out external cohorts, GSE40279 (n = 656) and GSE69270 (n = 184) were excluded from model fitting, leaving 8,986 samples for nested group-aware cross-validation. GSE207605 was therefore treated as a harmonized training resource rather than as a single independent external cohort.

GSE40279 was used as a classical independent whole-blood aging cohort for external comparison and age-stratified evaluation [2]. GSE141682 was included as the most directly forensic-relevant public blood dataset in this framework, comprising 42 healthy Han Chinese individuals aged 18–62 years [10]. GSE69270 was selected as an independent age-focused sensitivity cohort from the Young Finns Study, comprising 184 whole-blood samples analyzed on the Illumina HumanMethylation450 platform [11, 12].

A summary of all included cohorts is shown in Table 1, whereas fuller cohort annotations are provided in Online Resource 1, Table S1. Sample-level ages, cohort identifiers, and model predictions for the 9,826 complete cases are provided in Online Resource 2.

Table 1  
Public whole-blood DNA methylation datasets included in the final analytical framework

| GEO accession | Cohort / role                                | Platform          | Tissue      | Sample accounting                                                 | Age range (years) | Analytical use                                                                         |
|---------------|----------------------------------------------|-------------------|-------------|-------------------------------------------------------------------|-------------------|----------------------------------------------------------------------------------------|
| GSE207605     | Health-filtered harmonized training resource | 450K / integrated | Whole blood | 9,879 metadata; 9,831 feature-table records; 9,826 complete cases | 0-101             | Training resource; 8,986 samples fitted after excluding held-out GSE40279 and GSE69270 |
| GSE40279      | Whole-blood aging validation cohort          | 450K              | Whole blood | 656                                                               | 19–101            | Held-out external validation                                                           |
| GSE141682     | Healthy Han Chinese forensic cohort          | EPIC              | Whole blood | 42                                                                | 18–62             | Independent forensic-relevant external validation                                      |
| GSE69270      | Young Finns sensitivity cohort               | 450K              | Whole blood | 184                                                               | 40–49             | Held-out narrow-age-range sensitivity validation                                       |

### 3.4. Data retrieval and preprocessing

All public methylation data were downloaded from the corresponding GEO accession records. Where processed beta values were available, these were preferred to maximize reproducibility and reduce unnecessary preprocessing heterogeneity. For datasets that did not directly provide processed beta values, beta values were derived from the

publicly available array signal matrices using the corresponding methylated and unmethylated intensity information. Samples lacking usable chronological-age information were excluded.

Because cross-cohort integration can inflate apparent performance, cohort-level separation was maintained between model development and external evaluation whenever feasible [2, 3]. The 8,986-sample training set excluded GSE40279 and GSE69270 before fitting. GSE141682 was processed independently and used only for external validation. The sample-level analysis output is provided in Online Resource 2.

### **3.5. Forensic marker backbone and staged panel definitions**

A seed forensic marker backbone was defined from loci repeatedly reported in blood-based age-estimation studies. The backbone was not intended to be exhaustive but to provide a compact starting point for targeted assay design. ELOVL2, FHL2, PDE4C, KLF14, EDARADD, ASPA, MIR29B2CHG/C1orf132, and TRIM59 have recurrent support in the forensic literature [4–6, 13, 15, 17, 18]. Empirical model fitting was restricted to the nine CpGs available in the supplied feature table. Canonical KLF14 (cg14361627), ASPA (cg02228185), and MIR29B2CHG/C1orf132 (cg10501210) markers were therefore retained as literature-supported reference loci but were not used as model features. Marker-level literature support and empirical inclusion are summarized in Table 2 and Online Resource 1, Table S4A.

Table 2  
Main forensic age-associated markers and their operational roles within the staged panel framework

| Marker                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              | CpG        | Forensic support         | Zone                      | Operational role                                              | Empirical model status                           |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------|--------------------------|---------------------------|---------------------------------------------------------------|--------------------------------------------------|
| A. Empirically evaluated CpGs in the current study (n = 9)                                                                                                                                                                                                                                                                                                                                                                                                                                          |            |                          |                           |                                                               |                                                  |
| ELOVL2                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              | cg16867657 | Strong                   | Zone I                    | Key chronological-age core marker                             | Models A/B/C                                     |
| FHL2                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                | cg06639320 | Strong                   | Zone I                    | Compact-panel core marker                                     | Models A/B/C                                     |
| PDE4C                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               | cg17861230 | Strong                   | Zone I/II                 | Supports chronological-age backbone                           | Models A/B/C                                     |
| EDARADD                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             | cg09809672 | Strong                   | Zone I/II                 | Directional balance in compact prediction                     | Models A/B/C                                     |
| ELOVL2                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              | cg24724428 | Strong                   | Zone II                   | Second ELOVL2 site used for controlled expansion              | Models B/C                                       |
| FHL2                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                | cg22454769 | Strong                   | Zone II                   | Second FHL2 site used for controlled expansion                | Models B/C                                       |
| TRIM59                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              | cg07553761 | Moderate-strong          | Zone II                   | Interface-layer expansion                                     | Models B/C                                       |
| SLC12A5                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             | cg07547549 | Moderate-strong          | Zone II                   | Interface-layer expansion                                     | Models B/C                                       |
| Intergenic / no assigned gene                                                                                                                                                                                                                                                                                                                                                                                                                                                                       | cg17110586 | Age-associated candidate | Zone II/III (provisional) | Model C-only empirical extension; no biological role assigned | Model C                                          |
| B. Additional literature-supported forensic loci not included in the supplied nine-CpG feature table (n = 3)                                                                                                                                                                                                                                                                                                                                                                                        |            |                          |                           |                                                               |                                                  |
| KLF14                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               | cg14361627 | Strong                   | Zone I/II                 | Literature-supported forensic locus                           | Not included in the supplied 9-CpG feature table |
| ASPA                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                | cg02228185 | Moderate-strong          | Zone I/II                 | Literature-supported supporting locus                         | Not included in the supplied 9-CpG               |
| <b>Note:</b> Only the nine CpGs in Section A were included in the empirical staged models and subsequent statistical analyses. The three loci in Section B were retained solely as literature-supported forensic reference markers because their canonical CpGs were absent from the supplied nine-feature table. “Forensic support” is a qualitative summary of literature support, and the zone labels are operational panel-development categories rather than statistical significance classes. |            |                          |                           |                                                               |                                                  |

| Marker                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              | CpG        | Forensic support | Zone      | Operational role                    | Empirical model status                           |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------|------------------|-----------|-------------------------------------|--------------------------------------------------|
|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |            |                  |           |                                     | feature table                                    |
| MIR29B2CHG/C1orf132                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 | cg10501210 | Strong           | Zone I/II | Literature-supported forensic locus | Not included in the supplied 9-CpG feature table |
| <b>Note:</b> Only the nine CpGs in Section A were included in the empirical staged models and subsequent statistical analyses. The three loci in Section B were retained solely as literature-supported forensic reference markers because their canonical CpGs were absent from the supplied nine-feature table. “Forensic support” is a qualitative summary of literature support, and the zone labels are operational panel-development categories rather than statistical significance classes. |            |                  |           |                                     |                                                  |

## 3.6. Candidate marker evaluation

Candidate features within the supplied nine-CpG table were evaluated using a forensic panel-development strategy rather than a single ranking statistic. Considerations included chronological-age association in blood, directional consistency across available cohorts, prior forensic literature support, and compatibility with a compact targeted assay. Because the available analysis table contained nine candidate CpGs rather than genome-wide methylation features, this study should be interpreted as staged candidate-panel evaluation rather than exhaustive epigenome-wide marker discovery. Model definitions and marker annotations are provided in Online Resource 1, Tables S2 and S2A; cross-cohort CpG-age correlations are provided in Online Resource 1, Table S6.

## 3.7. Staged model definitions

Three Elastic Net model classes were compared. Model A was prespecified as a four-CpG literature-supported compact forensic core to represent the lowest-complexity panel compatible with targeted implementation [4–6, 13, 15, 17]. It was not selected as the data-driven optimum.

Models B and C were derived by training-set-internal stepwise expansion beyond Model A within the filtered GSE207605 training framework. Candidate additions were retained according to cross-validated performance within the staged panel-design logic. Model B contained eight CpGs and represented a controlled expansion of the compact core; Model C contained all nine available CpGs and represented the maximum empirical extension within the supplied candidate set. The comparison therefore tested whether modest increases in marker number produced practically meaningful gains rather than assuming that larger panels would perform better.

Elastic Net regression was selected because it provides regularized fitting and coefficient-level interpretability for correlated methylation predictors while remaining compatible with compact-panel development. Hyperparameters were selected within the inner group-aware folds, and final coefficients are reported in Online Resource 1, Table S4.

## 3.8. Model evaluation

Because forensic utility cannot be captured by a single global metric, model performance was evaluated using MAE, RMSE, Pearson correlation,  $R^2$ , mean prediction bias, age-stratified error, and cohort-wise transferability. Internal performance was estimated using nested group-aware cross-validation, with GEO accession defining groups in both outer and inner folds. GSE40279 and GSE69270 were removed from model fitting and evaluated as

held-out cohorts; GSE141682 was independently processed from its EPIC series matrix and used as a third external cohort. Predictions were retained on the original linear scale for transparent calibration assessment.

A calibration sensitivity analysis used 1,000 repeated rank-age-stratified splits within each external cohort. In every split, 30% of samples were used only to estimate calibration parameters and the remaining 70% were retained as untouched test samples. The primary method adjusted only the intercept by adding the mean calibration-subset residual to each test prediction. A secondary model estimated both intercept and slope, and the calibration-subset mean age was evaluated as a local baseline. Raw external predictions remained the primary validation endpoint.

For future wet-lab translation, additional analytical criteria should include DNA input requirements, tolerance to sample degradation, repeatability, reproducibility, and compatibility with targeted bisulfite-based workflows [4–6, 13–15]. These criteria were not experimentally evaluated in the present public-data study and should be addressed prospectively before operational forensic use.

## 4. Results

### 4.1. Forensic marker backbone and cross-cohort age associations

A consistent finding from the forensic literature is convergence on a relatively small set of informative blood-associated loci. ELOVL2 is the most consistently reported, while FHL2, PDE4C, KLF14, EDARADD, ASPA, MIR29B2CHG/C1orf132, and TRIM59 recur across compact forensic panels [4–6, 13, 15, 17, 18]. The empirical staged feature table available here represented a subset of that literature backbone. Across GSE40279, GSE141682, and GSE69270, cg16867657 and cg24724428 (both ELOVL2) showed strong positive age correlations; cg06639320 and cg22454769 (both FHL2) were also consistently positive. TRIM59 (cg07553761) and SLC12A5 (cg07547549) showed positive associations, whereas EDARADD (cg09809672) was negative. PDE4C (cg17861230) was positive but more cohort dependent, particularly in the narrow-age-range GSE69270 cohort. Canonical KLF14, ASPA, and MIR29B2CHG/C1orf132 CpGs were not present in the supplied nine-feature modeling table and therefore were not interpreted as empirical model features in this analysis.

From a forensic perspective, the importance of these findings lies in the convergence of literature support with reproducible age associations among the staged-model CpGs. The main literature-supported backbone and empirical model status are summarized in Table 2 and Online Resource 1, Table S4A. CpG–age correlations for the nine staged-model markers in the three external cohorts are reported in Online Resource 1, Table S6. Online Resource 1, Fig. S2 visualizes cross-cohort correlation patterns across the health-filtered harmonized resource, and Online Resource 1, Fig. S1 summarizes cohort sample sizes and observed age ranges.

### 4.2. Training resource and candidate-panel scope

GSE207605 provided the primary multi-cohort training resource for staged model development. After complete-case filtering and removal of samples from GSE40279 and GSE69270, 8,986 samples remained for nested group-aware cross-validation. GSE141682 was processed independently and used only for external evaluation. The resulting design preserved separation between model development and all three external evaluations.

Because the supplied model-feature table contained only nine candidate CpGs, the analysis evaluates a predefined candidate space rather than genome-wide discovery. This distinction is important for interpreting the comparison

between Models A-C: the study tests whether controlled expansion of a compact forensic panel is useful within the available features, not whether these nine CpGs are globally optimal among all methylation sites.

### **4.3. Controlled marker expansion beyond the compact core**

Model B added a second ELOVL2 CpG (cg24724428), a second FHL2 CpG (cg22454769), TRIM59 (cg07553761), and SLC12A5 (cg07547549) to the four-CpG core. Model C further added cg17110586, an intergenic/unassigned age-associated CpG. This design allowed incremental evaluation of marker expansion while retaining a panel size compatible with future targeted workflows.

The expanded markers showed age associations across the external cohorts, but the strength of association varied by cohort and age range. The staged comparison was therefore interpreted in terms of both prediction error and transferability rather than marker count alone. The exact model compositions are provided in Online Resource 1, Table S2.

### **4.4. Staged panel architecture for forensic evaluation**

Model A represented the prespecified compact core; Model B represented the controlled eight-CpG expansion; and Model C represented the nine-CpG extended panel. This architecture provided a direct test of whether additional CpGs improved chronological-age prediction under cohort-aware validation.

Internal cross-validation favored the expanded models only modestly, whereas external ranking varied by cohort. The practical question was therefore not whether expansion could reduce internal error, but whether any improvement was sufficiently consistent to justify additional markers in a targeted forensic assay.

The overall workflow is shown in Fig. 1, and the conceptual relationship between compact forensic prediction, controlled expansion, and broader aging interpretation is shown in Fig. 2. Exact model specifications are provided in Online Resource 1, Table S2.

### **4.5. Operational interpretation of panel size**

Across the staged models, increasing marker number did not produce uniform improvement in external accuracy. This finding supports treating compactness as an explicit forensic design objective rather than assuming that a larger methylation panel is intrinsically preferable. Controlled expansion may still be useful when it improves robustness or calibration behavior, but its value must be demonstrated across independent cohorts.

Accordingly, the four-CpG core was retained as the primary low-complexity benchmark, while Models B and C were interpreted as tests of incremental expansion. Broader biological-age interpretation was not used to rank the forensic models and remains outside the primary casework endpoint.

### **4.6. Comparative performance of staged panels**

Chronological-age prediction was compared across the three panel sizes under nested group-aware internal validation and three independent external evaluations. Model A contained four CpGs, Model B eight, and Model C nine.

Under nested group-aware cross-validation of the health-filtered GSE207605 training resource (8,986 samples; GEO accession used as the grouping unit), Model A achieved an MAE of 5.97 years, an RMSE of 7.65 years, a Pearson correlation of 0.938, and an  $R^2$  of 0.877. Model B improved MAE to 5.86 years and RMSE to 7.23 years ( $r = 0.946$ ;  $R^2$

= 0.890), whereas Model C yielded only a marginal MAE improvement (5.84 years) and a slightly higher RMSE (7.30 years;  $r = 0.944$ ;  $R^2 = 0.888$ ).

External validation demonstrated substantial cohort dependence. In GSE40279, Model A yielded an MAE of 8.62 years and an RMSE of 10.42 years; Model C produced similar RMSE (10.42 years) but a slightly higher MAE (8.68 years), while all models showed systematic underestimation (bias approximately - 7.4 to -7.8 years). In the healthy Han Chinese cohort GSE141682, Model B performed best (MAE 7.85 years; RMSE 8.88 years;  $r = 0.965$ ;  $R^2 = 0.710$ ), although all models showed positive calibration bias. In the narrow-age-range GSE69270 cohort, Model A retained the lowest error (MAE 3.90 years; RMSE 4.62 years), but all models had negative  $R^2$  values because between-person age variance was limited. These results support controlled expansion while showing that a larger marker set is not uniformly superior across cohorts.

Taken together, these findings support the practical value of staged model development, while emphasizing that raw external model ranking was cohort dependent. Expansion beyond the compact forensic core modestly improved internal cross-validation performance but did not consistently improve raw external accuracy. Model A performed best in GSE40279 and GSE69270, whereas Model B showed a marginal advantage in GSE141682. Thus, the comparison supports a staged strategy in which a minimal forensic core remains primary, limited expansion is evaluated for robustness and calibration behavior, and further extension should not be interpreted as automatically preferable.

## 4.7. Leakage-free external calibration sensitivity analysis

The raw external results showed strong cohort-dependent offset despite preserved age correlation. Under 1,000 repeated 30%/70% calibration/test splits, intercept-only calibration reduced Model B mean held-out MAE from 8.83 to 5.56 years in GSE40279, from 4.28 to 2.46 years in GSE69270, and from 7.77 to 3.92 years in GSE141682. The corresponding reductions were 37.1%, 42.5%, and 49.5%, and the calibrated MAE was lower in all repeated splits. Mean held-out bias was reduced to approximately zero.

Two-parameter linear calibration produced slightly lower MAE in some settings, but was treated as a secondary sensitivity analysis because slope estimation is less stable with small calibration subsets. Intercept-only calibration was therefore selected as the conservative calibration analysis. These estimates do not replace the raw independent validation; they quantify the potential value of local laboratory reference samples.

Table 3  
Leakage-free intercept-only calibration sensitivity analysis

| Cohort    | Model | Calibration/Test N | Mean uncalibrated held-out MAE | Mean calibrated held-out MAE (95% interval) | Reduction | Calibrated RMSE | Calibrated bias |
|-----------|-------|--------------------|--------------------------------|---------------------------------------------|-----------|-----------------|-----------------|
| GSE40279  | A     | 196/460            | 8.62                           | 5.84 (5.66–6.03)                            | 32.2%     | 7.33            | -0.04           |
| GSE40279  | B     | 196/460            | 8.83                           | 5.56 (5.36–5.75)                            | 37.1%     | 7.09            | -0.03           |
| GSE40279  | C     | 196/460            | 8.69                           | 5.56 (5.37–5.75)                            | 36.0%     | 7.09            | -0.03           |
| GSE69270  | A     | 55/129             | 3.89                           | 2.54 (2.36–2.71)                            | 34.9%     | 3.14            | -0.02           |
| GSE69270  | B     | 55/129             | 4.28                           | 2.46 (2.28–2.65)                            | 42.5%     | 3.09            | -0.02           |
| GSE69270  | C     | 55/129             | 4.46                           | 2.47 (2.30–2.67)                            | 44.6%     | 3.11            | -0.02           |
| GSE141682 | A     | 12/30              | 7.86                           | 4.23 (3.68–4.83)                            | 46.2%     | 5.13            | -0.39           |
| GSE141682 | B     | 12/30              | 7.77                           | 3.92 (3.41–4.49)                            | 49.5%     | 4.79            | -0.32           |
| GSE141682 | C     | 12/30              | 8.00                           | 4.02 (3.50–4.60)                            | 49.8%     | 4.87            | -0.31           |

## 5. Discussion

This study evaluated a practical problem in forensic epigenetic age estimation: whether a compact, literature-supported blood methylation panel remains competitive when modestly expanded, and how much of the external prediction error reflects cohort-specific calibration rather than loss of age signal. Three findings are central. First, expansion from four to eight or nine CpGs produced only small gains in nested group-aware internal validation. Second, no expanded model was uniformly superior across external cohorts. Third, a simple leakage-free intercept correction substantially reduced held-out MAE in all three external datasets. Together, these results place transferability and calibration, rather than marker number alone, at the center of forensic panel development.

The four-CpG Model A was deliberately defined as a low-complexity benchmark based on repeatedly supported forensic loci. Its internal MAE of 5.97 years was only 0.11–0.13 years higher than Models B and C, and it achieved the lowest raw MAE in GSE40279 and GSE69270. Model B showed a small advantage in GSE141682, but the

differences among models were modest relative to the between-cohort shifts in absolute error. For a targeted forensic assay, this trade-off matters because every added CpG increases assay design, validation, and quality-control burden. The results therefore do not argue against panel expansion; rather, they indicate that added markers should be justified by reproducible external benefit.

The external results also illustrate why correlation alone is insufficient for forensic validation. GSE40279 and GSE141682 retained strong age-prediction correlations while showing systematic under- or overestimation, respectively. GSE69270 had a restricted 40-49-year age range and negative  $R^2$  values despite measurable within-cohort correlation, demonstrating how range restriction can distort variance-based performance metrics. Reporting MAE, RMSE, bias, correlation,  $R^2$ , and age-stratified error together therefore provides a more informative picture than any single statistic.

The calibration sensitivity analysis provides the most practically relevant extension of the raw external validation. Using only the 30% calibration subset in each repeated split, intercept-only correction reduced Model B mean held-out MAE from 8.83 to 5.56 years in GSE40279, from 4.28 to 2.46 years in GSE69270, and from 7.77 to 3.92 years in GSE141682. Because the untouched 70% test subset was never used to estimate the correction, these results avoid direct calibration leakage. The analysis does not establish universal portability, but it shows that cohort-specific offset can account for a substantial proportion of absolute error when local reference samples are available.

This interpretation is consistent with recent work emphasizing assay translation rather than prediction accuracy in isolation. Interval-based ELOVL2 modeling explicitly represents uncertainty [20], a three-marker droplet digital PCR approach demonstrates the feasibility of targeted quantitative workflows [21], population-specific modeling highlights ancestry and calibration effects [22], and a multi-laboratory exercise shows that standardization can support cross-laboratory implementation [23]. GP-age demonstrates the value of large cross-cohort resources for portable age prediction, but its 30-CpG architecture addresses a different operational balance from a minimal forensic panel [24]. The present results add to this literature by directly comparing constrained panel sizes under cohort-aware validation and separating raw external performance from local calibration sensitivity.

The restriction to whole blood is both a strength and a limitation. It reduces tissue-related heterogeneity and keeps the analysis relevant to one of the most extensively studied matrices in forensic methylation research. At the same time, it does not establish performance in saliva, buccal cells, semen, bone, teeth, nails, or other forensic materials [4–6, 14, 15]. Even within blood, platform differences, preprocessing history, cell composition, ancestry, and age distribution may contribute to systematic offset. Future validation should therefore treat population and laboratory calibration as explicit components of assay qualification rather than post hoc adjustments.

The broader biological-aging literature was included only to provide context for age-associated methylation signals outside the minimal forensic core. The present study does not validate a biological-age clock and does not support incorporating biological-age deviation into routine forensic reporting. Chronological age remains the primary forensic endpoint. Keeping these objectives separate is important because a marker set optimized for health-related aging phenotypes may not be optimal for casework accuracy, assay simplicity, or interpretive clarity.

Several design features strengthen the analysis. Overlapping GSE40279 and GSE69270 samples were removed from the GSE207605-derived training resource before model fitting, group-aware nested cross-validation used GEO accession as the grouping unit, GSE141682 was independently processed as a forensic-relevant external cohort, and calibration parameters were estimated only in dedicated calibration subsets. In addition, the study reports raw

external estimates as the primary validation results and treats calibration as a sensitivity analysis rather than replacing independent validation.

Important limitations remain. This is a public-data secondary analysis rather than a de novo laboratory validation study. The candidate space was limited to nine supplied CpGs, so the results cannot establish genome-wide optimality or exclude better marker combinations. Public datasets also differ in platform, preprocessing, cohort structure, and available metadata. GSE141682 is small, and its repeated-split calibration subsets contain only 12 samples. The 30% calibration requirement may be impractical in some forensic settings and should be evaluated prospectively.

The modeling strategy was intentionally conservative. Elastic Net was prioritized over higher-capacity machine-learning methods because the scientific question concerned staged panel architecture, interpretability, and transferability rather than unconstrained predictive optimization. Models B and C were selected within the training framework, so some internal optimism cannot be excluded despite nested group-aware validation. For that reason, the external results are more important than the small differences observed internally.

Before casework application, the candidate panels require prospective wet-lab evaluation using targeted methylation assays and practice-relevant specimens. Key studies should quantify DNA input requirements, tolerance to degradation, repeatability, reproducibility, bisulfite-conversion effects, inter-laboratory variation, and performance across populations and age ranges. Calibration procedures should be prespecified and tested with realistically sized local reference sets. Interval or uncertainty-based reporting may also be preferable to presenting a single age estimate without context [20].

Overall, the data support a restrained development strategy: begin with a compact forensic core, add markers only when they deliver reproducible external benefit, and treat local calibration as a formal component of transferability assessment. In this framework, a slightly more accurate internal model is not automatically a better forensic assay. The central criterion is robust, interpretable performance under the conditions in which the method will actually be used.

## 6. Conclusion

In public whole-blood DNA methylation data, a four-CpG forensic core remained competitive with eight- and nine-CpG expansions despite small internal gains from additional markers. External model ranking was cohort dependent, indicating that increasing marker number alone does not ensure better transferability.

Leakage-free intercept calibration substantially reduced held-out error in all three external cohorts, supporting cohort- and laboratory-specific offset as an important consideration in forensic age estimation. Raw independent validation remains the primary evidence, and calibration should be prespecified rather than used to mask poor transportability.

These findings support a staged route to implementation: compact panel first, controlled expansion only when externally justified, and prospective wet-lab validation under low-input, degraded, population-specific, and inter-laboratory conditions before forensic casework use.

## Declarations

### Ethics approval

Not applicable. This secondary analysis used only publicly available, de-identified datasets and did not involve recruitment of new participants or collection of new specimens.

#### **Clinical trial number**

Not applicable.

#### **Consent to participate and publication**

Not applicable.

#### **Data availability**

All datasets analyzed in this study are publicly available through the Gene Expression Omnibus (GEO). No new raw data were generated. Sample-level analysis output is provided as Online Resource 2.

#### **Competing interests**

The authors declare that they have no competing interests.

#### **Funding**

No funding was received for conducting this study.

## **Author Contribution**

Baihereye Abudouwaier: Data curation, Writing - original draft. Shengjie Gao: Conceptualization, Methodology, Writing - review and editing. Halimureti Simayijiang: Conceptualization, Formal analysis, Supervision, Writing - review and editing. All authors read and approved the final manuscript.

## **Acknowledgement**

The authors thank the investigators who generated and shared the public DNA methylation datasets used in this study.

## **References**

1. Horvath S (2013) DNA methylation age of human tissues and cell types. *Genome Biol* 14:R115. <https://doi.org/10.1186/gb-2013-14-10-r115>
2. Hannum G, Guinney J, Zhao L et al (2013) Genome-wide methylation profiles reveal quantitative views of human aging rates. *Mol Cell* 49(2):359–367. <https://doi.org/10.1016/j.molcel.2012.10.016>
3. Field AE, Robertson NA, Wang T, Havas A, Ideker T, Adams PD (2018) DNA methylation clocks in aging: categories, causes, and consequences. *Mol Cell* 71(6):882–895. <https://doi.org/10.1016/j.molcel.2018.08.008>
4. Huang Y, Yan J, Hou J et al (2015) Developing a DNA methylation assay for human age prediction in blood and bloodstain. *Forensic Sci International: Genet* 17:129–136. <https://doi.org/10.1016/j.fsigen.2015.05.007>

5. Jung SE, Lim SM, Hong SR, Lee EH, Shin KJ, Lee HY (2019) DNA methylation of the ELOVL2, FHL2, KLF14, C1orf132/MIR29B2C, and TRIM59 genes for age prediction from blood, saliva, and buccal swab samples. *Forensic Sci International: Genet* 38:1–8. <https://doi.org/10.1016/j.fsigen.2018.09.010>
6. Correia Dias H, Cordeiro C, Corte Real F, Cunha E, Manco L (2020) Age estimation based on DNA methylation using blood samples from deceased individuals. *J Forensic Sci* 65(2):465–470. <https://doi.org/10.1111/1556-4029.14185>
7. Levine ME, Lu AT, Quach A et al (2018) An epigenetic biomarker of aging for lifespan and healthspan. *Aging* 10(4):573–591. <https://doi.org/10.18632/aging.101414>
8. Lu AT, Quach A, Wilson JG et al (2019) DNA methylation GrimAge strongly predicts lifespan and healthspan. *Aging* 11(2):303–327. <https://doi.org/10.18632/aging.101684>
9. NCBI Gene Expression Omnibus GSE207605: Accurate age prediction from blood using a small set of DNA methylation sites and a cohort-based machine learning algorithm. GEO DataSets. ([ncbi.nlm.nih.gov](https://ncbi.nlm.nih.gov))
10. NCBI Gene Expression Omnibus GSE141682: DNA methylation profiling of whole blood from Han Chinese individuals: identification of age-related CpG sites for forensic use. GEO DataSets. ([ncbi.nlm.nih.gov](https://ncbi.nlm.nih.gov))
11. NCBI Gene Expression Omnibus GSE69270: Aging-associated DNA methylation changes in middle-aged individuals: The Young Finns Study. GEO DataSets. ([ncbi.nlm.nih.gov](https://ncbi.nlm.nih.gov))
12. Kananen L, Marttila S, Nevalainen T et al (2016) Aging-associated DNA methylation changes in middle-aged individuals: the Young Finns study. *BMC Genomics* 17:103. <https://doi.org/10.1186/s12864-016-2421-z>
13. Zbieć-Piekarska R, Spólnicka M, Kupiec T, Parys-Proszek A, Makowska Ż, Pałeczka A et al (2015) Development of a forensically useful age prediction method based on DNA methylation analysis. *Forensic Sci Int Genet* 17:173–179. <https://doi.org/10.1016/j.fsigen.2015.05.001>
14. Hong SR, Jung SE, Lee EH, Shin KJ, Yang WI, Lee HY (2017) DNA methylation-based age prediction from saliva: High age predictability by combination of 7 CpG markers. *Forensic Sci Int Genet* 29:118–125. <https://doi.org/10.1016/j.fsigen.2017.04.006>
15. Bekaert B, Kamalandua A, Zapico SC, Van de Voorde W, Decorte R (2015) Improved age determination of blood and teeth samples using a selected set of DNA methylation markers. *Epigenetics* 10(10):922–930. <https://doi.org/10.1080/15592294.2015.1080413>
16. Daunay A, Baudrin LG, Deleuze JF, How-Kit A (2019) Evaluation of six blood-based age prediction models using DNA methylation analysis by pyrosequencing. *Sci Rep* 9:8862. <https://doi.org/10.1038/s41598-019-45197-w>
17. Weidner CI, Lin Q, Koch CM, Eisele L, Beier F, Ziegler P et al (2014) Aging of blood can be tracked by DNA methylation changes at just three CpG sites. *Genome Biol* 15(2):R24. <https://doi.org/10.1186/gb-2014-15-2-r24>
18. Garagnani P, Bacalini MG, Pirazzini C, Gori D, Giuliani C, Mari D et al (2012) Methylation of ELOVL2 gene as a new epigenetic marker of age. *Aging Cell* 11(6):1132–1134. <https://doi.org/10.1111/acer.12005>
19. Bocklandt S, Lin W, Sehl ME, Sánchez FJ, Sinsheimer JS, Horvath S et al (2011) Epigenetic predictor of age. *PLoS ONE* 6(6):e14821. <https://doi.org/10.1371/journal.pone.0014821>
20. Yamagishi T, Sakurai W, Watanabe K, Toyomane K, Akutsu T (2024) Development and comparison of forensic interval age prediction models by statistical and machine learning methods based on the methylation rates of ELOVL2 in blood DNA. *Forensic Sci Int Genet* 69:103004. <https://doi.org/10.1016/j.fsigen.2023.103004>
21. Correia Dias H, Manco L (2024) Predicting age from blood by droplet digital PCR using a set of three DNA methylation markers. *Forensic Sci Int* 356:111950. <https://doi.org/10.1016/j.forsciint.2024.111950>

22. Filoglu G, Simsek SZ, Ersoy G, Can K, Bulbul O (2024) Epigenetic-based age prediction in blood samples: Model development. J Forensic Sci 69(3):869–879. <https://doi.org/10.1111/1556-4029.15478>

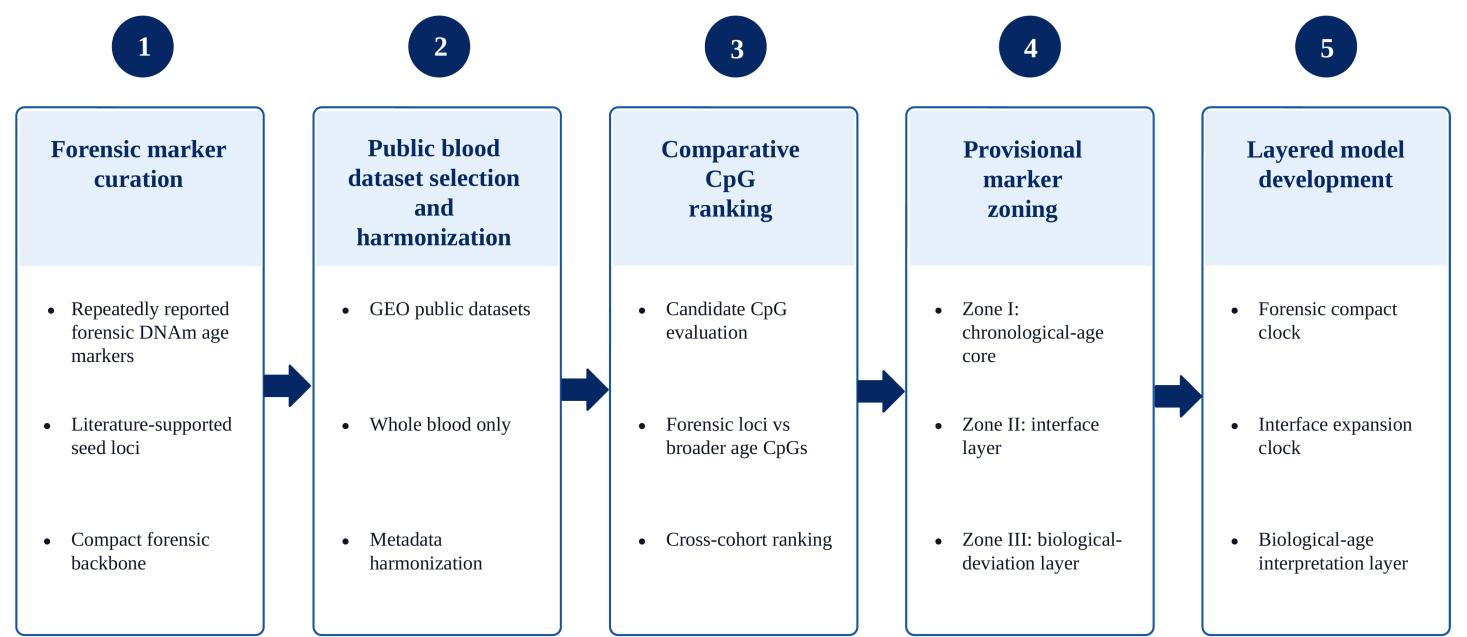
23. Kampmann ML, Fleckhaus J, Børsting C, Jurtikova H, Piters A, Papin J et al (2024) Collaborative exercise: analysis of age estimation using a QIAGEN protocol and the PyroMark Q48 platform. Forensic Sci Res 9(1):owad055. <https://doi.org/10.1093/fsr/owad055>

24. Varshavsky M, Harari G, Glaser B, Dor Y, Shemer R, Kaplan T (2023) Accurate age prediction from blood using a small set of DNA methylation sites and a cohort-based machine learning algorithm. Cell Rep Methods 3(9):100567. <https://doi.org/10.1016/j.crmeth.2023.100567>

Figures

Figure 1. Overall study design and analytical workflow

Forensic-oriented public-data analysis with a translational panel-development objective



Workflow combines literature curation, public-data harmonization, comparative ranking, operational zoning, and layered model development.

Figure 1

Overall study design and analytical workflow for compact forensic panel evaluation across public whole-blood DNA methylation datasets

## Figure 2. Conceptual layered bridge model

*Linking forensic chronological age estimation and biological age assessment*

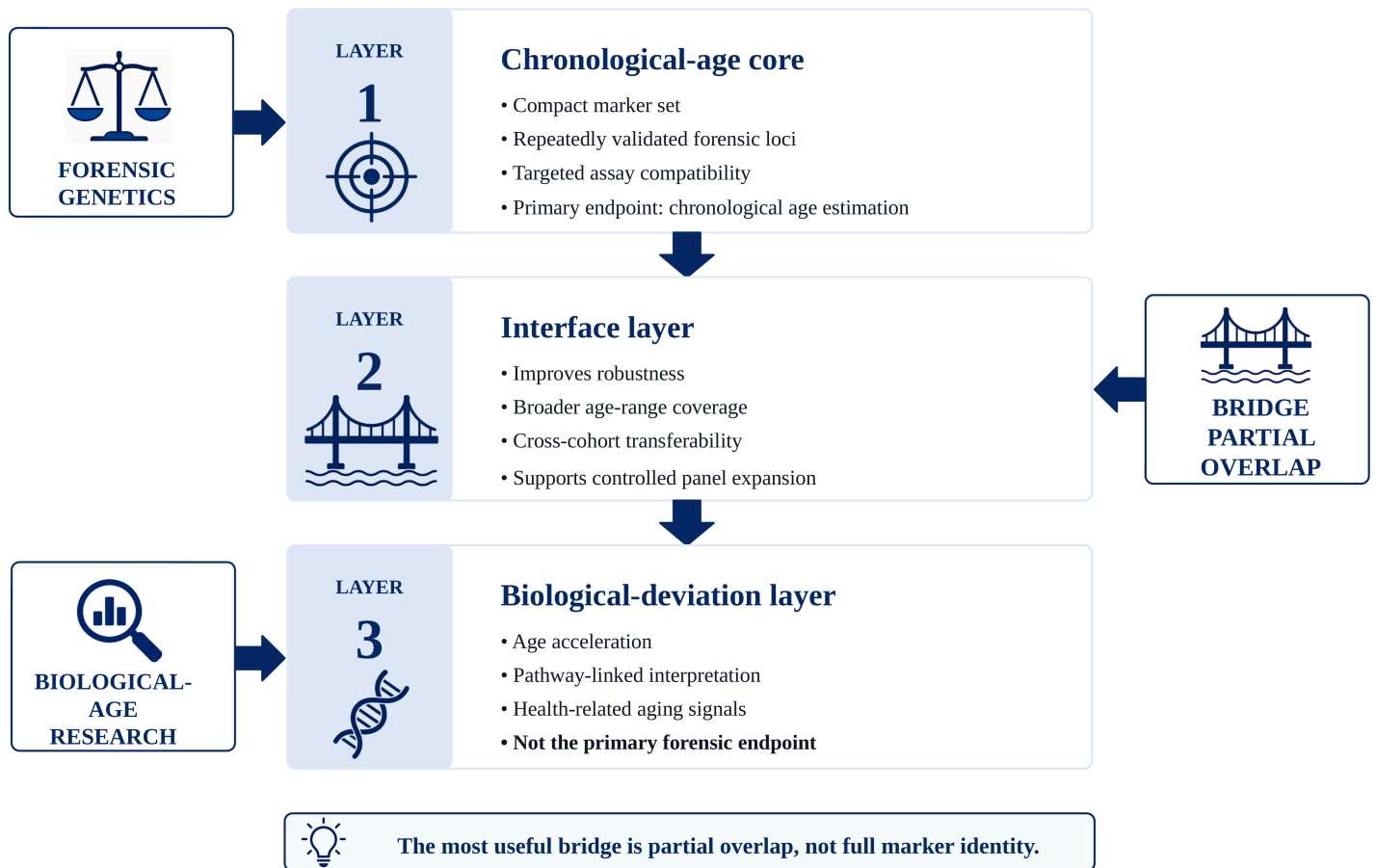


Figure 2

Conceptual context for staged forensic panel development linking a compact chronological-age core to controlled marker expansion; broader biological-age interpretation remains secondary to the forensic endpoint

## Staged methylation models under cohort-aware validation

A. Error metrics



B. Observed vs predicted age — Model B

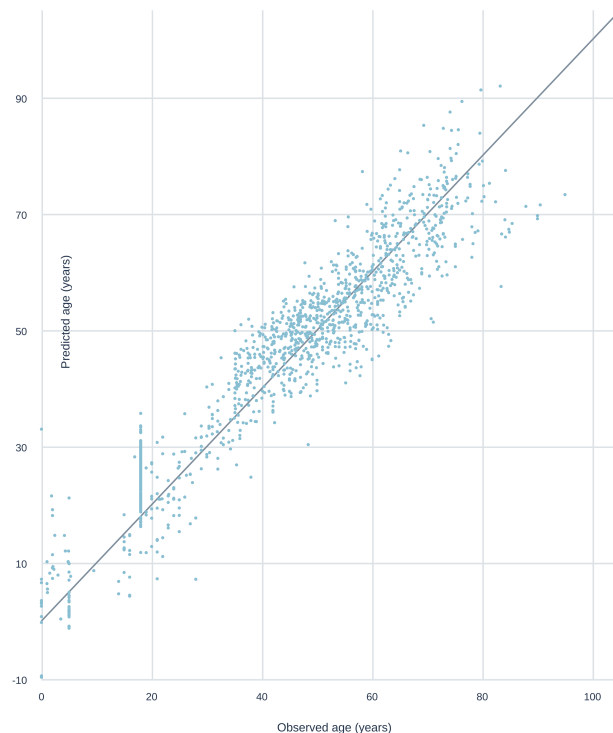


Figure 3

**Performance of staged methylation models under cohort-aware validation. Panel a compares mean absolute error across internal group-aware cross-validation and the two larger held-out cohorts (GSE40279 and GSE69270) shown in the primary comparison; complete external results including GSE141682 are reported in Online Resource 1, Table S3. Panel b shows observed versus predicted age for Model B in the group-aware internal evaluation**

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [OnlineResource1SupplementaryMaterials.docx](#)
- [OnlineResource2SupplementaryDataS1.xlsx](#)