

Tissue-specific Mitotic Clocks Link Cumulative Stem Cell Divisions to Human Cancer Susceptibility

Fredrick Schumacher

frs2@case.edu

Case Western Reserve University <https://orcid.org/0000-0002-3073-7463>

Yanning Wu

Case Western Reserve University

Article

Keywords:

Posted Date: March 9th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9013656/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: There is **NO** Competing Interest.

Abstract

Why cancer risk varies so dramatically across human tissues remains a fundamental question in oncology. While this variation has been linked to differences in total stem cell divisions (TNSD), molecular tools for measuring tissue-specific proliferative history are lacking. Existing DNA methylation-based mitotic clocks conflate chronological aging with mitotic activity and apply pan-tissue assumptions that obscure tissue-level turnover dynamics. Here, we develop tissue-specific mitotic clocks for six major tissues by modeling global hypomethylation at solo-WCGW CpG sites within partially methylated domains (PMDs) – an experimentally validated, age-independent marker of cumulative cell division – using biologically normal tissue from the Genotype-Tissue Expression (GTEx) project. This design explicitly avoids confounding from chronological age and field cancerization. Our tissue-specific clocks demonstrate strong generalizability, replicate across independent cohorts, and outperform existing pan-tissue clocks in recapitulating Vogelstein and Tomasetti-derived (2017) TNSD estimates and SEER cancer incidence data. These results provide direct molecular evidence that cumulative cellular turnover underlies tissue cancer susceptibility, and establish a precision framework for quantifying proliferative burden in human malignancies – with implications for cancer risk stratification, prevention, and mechanistic dissection of replication-driven versus exposure-related carcinogenesis.

Introduction

Accumulating evidence indicates that the variation of lifetime cancer risk across human tissues is strongly associated with the total number of stem cell divisions (TNSD) that occur within each tissue over a lifetime^{1–5}. Among the molecular alterations that arise during cellular proliferation, DNA methylation (DNAm) changes – resulting from imperfect maintenance during DNA replication – accumulate with successive stem cell divisions, capturing both the biological progression and proliferative history of a tissue, and may ultimately predispose certain cell populations to the potential of malignant transformation^{6–9}.

To quantify this process, several DNAm-based mitotic clocks have been developed to estimate TNSD using distinct modeling strategies. Some clocks are tailored to specific tissues or cancer types. For example, Duran-Ferrer *et al.* introduced epiCMIT, a mitotic clock specifically designed to reflect the proliferative history of B-cell tumors, using both hyper- and hypomethylated CpGs in silent chromatin regions¹⁰. Other clocks target broader application across tissues. The MiAge calculator (Wang *et al.*) applied a statistical model based on methylation transmission theory to tumor and adjacent normal data using publicly-available data from The Cancer Genome Atlas (TCGA) program sponsored by the National Cancer Institute, identifying increasing-methylated probes that are informative of mitotic activity across tissues¹¹. A research group led by Andrew Teschendorff has developed a series of mitotic clocks – EpiTOC1¹², EpiTOC2¹³, and StemTOC¹⁴ – primarily derived from whole blood DNAm data. EpiTOC1 focused on CpG sites located within the promoter regions of Polycomb group target (PCGT) genes that are unmethylated in fetal tissues but progressively gain methylation with chronological age. EpiTOC2

refined this model by introducing mathematical formulations to estimate TNSD based on CpGs from EpiTOC1. StemTOC further extended this framework by implementing a multi-step filtering procedure that integrated DNAm data from fetal tissues, in-vitro cell lines, and in-vivo whole blood samples.

Most established DNAm-based mitotic clocks relied on a core assumption: chronological age serves as a direct surrogate for mitotic age. This assumption is applied in multiple ways – either explicitly, by regressing CpGs methylation levels against chronological age to identify mitotically informative CpGs during model development, or the implicit validation benchmark where a strong correlation with chronological age is taken as evidence of model credibility. However, this general assumption warrants closer scrutiny: biological aging (e.g. cumulative stem cell divisions^{15,16}) is unlikely follow a simple linear relationship with chronological age, as the rate of cellular turnover varies across life stages such as puberty and menopause, and is further influenced by many factors including inflammation¹⁷, oxidative stress¹⁸, and microenvironment¹⁹. In contrast, chronological age reflects merely the passage of time. Thus, these clocks may, in fact, better predict chronological aging but remain limited in capturing the complex biological processes that underpin mitotic activity.

Advancing beyond chronological age, we propose that global partially methylated domains (PMDs) hypomethylation offer a more precise surrogate of TNSD. Laird *et al.* first characterized PMDs using whole-genome bisulfite sequencing of TCGA adjacent normal tissues, identifying widespread hypomethylation specifically at solo-WCGW CpGs – isolated CpG sites occurring in an A/T – C – G – A/T (WCGW) sequence context – within PMDs²⁰. They observed that these sites are initially hypermethylated but progressively lose methylation as cells divide, suggesting that methylation loss at PMD solo-WCGWs reflects TNSD rather than the passage of time²⁰. A subsequent study from the same group provided experimental evidence using primary human cell cultures, showing that loss of DNA methylation at PMD solo-WCGWs tracks cumulative population doublings²¹. Collectively, these findings demonstrated that global hypomethylation at PMD solo-WCGW sites (global PMD hypomethylation) represents a biologically grounded and experimentally validated proxy for TNSD, offering a more accurate measure of proliferative history compared to chronological age.

However, global PMD hypomethylation identified by Laird *et al.* relied on a single adjacent normal sample per tissue type, which may not adequately capture tissue-specific methylation variation. In addition, several studies have reported tissue-specific differences in DNAm-based aging and emphasized the importance of developing tissue-specific biomarkers of aging^{22–24}. Therefore, we aim to build and validate tissue-specific mitotic clocks to estimate TNSD for six common tissue types. Unlike many existing models based on TCGA adjacent “normal” tissues, which potentially are impacted by field effects^{25–27} or pre-neoplastic epigenetic alterations²⁸, we used Genotype-Tissue Expression (GTEx) normal tissue to minimize these potential biases and enhance biological validity.

Results

Descriptive statistics of GTEx data

DNA methylation (DNAm) data were generated for 836 tissue samples obtained from 421 donors in the GTEx cohort, spanning six tissue types. The characteristics of GTEx donors for each tissue type are summarized in Table 1. Each donor contributed from 1 to 6 tissue samples. Donor ages ranged from 20 to 70 years, with a mean age of 54.4 ± 12.1 years; notably, approximately 90% of donors across all tissue types were aged over 40, indicating the samples used for developing tissue-specific mitotic clocks were biologically mature. For non-sex-specific tissues, more than 70% of samples were derived from male donors. The cohort was predominantly White (86.1%), with a mean BMI of 27.1 ± 4.0 kg/m². Additionally, 70.1% and 77.2% of donors reported smoking and alcohol use, respectively. Missing data across demographic and lifestyle variables were minimal (< 1%).

Tissue-specific methylation variation within PMDs

Larid *et al.* identified 26,732 PMD solo-WCGW CpGs represented on the EPIC array, of which 23,328 (87.2%) passed quality control (QC) in GTEx data and were retained for downstream analysis. The distribution of probes that did not pass QC showed no chromosome-specific bias (Supplementary Table 2). Using methylation levels at these CpGs, we performed PCA across 836 tissue samples from 421 GTEx donors to assess tissue-level methylation structure. As shown in Fig. 1, the first principal component (PC1) clearly separates ovarian samples from other tissues, while the remaining five tissue types form well-defined clusters. This pattern indicates substantial tissue-specific methylation variation within PMDs at the molecular level. However, the second (PC2) and third (PC3) principal components are non-informative.

To determine whether the observed tissue clustering was driven by inter-individual differences or chronological age-related effects, we repeated the PCA using matched multi-tissue samples from the same individuals. A total of 10 donors contributed matched breast, colon and ovary samples. As shown in Supplementary Figure S1, distinct tissue clusters were preserved. In contrast, comparison of global PMD hypomethylation across these 3 matched tissues showed no statistically significant differences (*P*-value = 0.089; Supplementary Figure S2). Together, these results demonstrate tissue-specific methylation variation at PMD solo-WCGW CpGs, likely reflecting distinct tissue-specific mitotic aging dynamics that are not detectable using global PMD methylation measures, thereby providing strong evidence for the necessity of developing tissue-specific mitotic clocks.

Construction and validation of tissue-specific mitotic clocks

We hypothesized that global PMD hypomethylation represents a more precise surrogate for TNSD than chronological age. Accordingly, we trained tissue-specific mitotic clocks for each tissue type using GTEx data (Methods). For each tissue type, a subset of informative CpGs and corresponding coefficient was identified (Supplementary Table 1). For a given sample, the tissue-specific mitotic clock generates a

tissue-specific hypomethylation score, by summing the weighted contributions of tissue-specific CpGs using their corresponding model coefficients. Lower scores indicate a higher degree of global PMD hypomethylation and therefore correspond to a higher estimated TNSD.

Next, we evaluated generalizability of the tissue-specific mitotic clocks using TCGA adjacent normal samples as an independent external validation cohort for the corresponding six tissue types (Table 2). Across tissues, we observed a consistent systematic bias in model estimates, ranging from -0.193 in ovary to -0.359 in kidney. RMSE values ranged from 0.19 to 0.359, reflecting overall deviation of predictions from observed values. Given the 0–1 scale of tissue-specific mitotic clock estimates, an RMSE of ~ 0.3 indicates moderate deviations on average, suggesting that overall TNSD trends are well captured.

Table 2
External validation of tissue-specific mitotic clocks on TCGA adjacent normal samples

Tissue	Number of Samples	Array	Performance metrics	
			Bias	RMSE
Prostate	50	HM450k	-0.276	0.28
Breast	97	HM450k	-0.358	0.358
Lung	74	HM450k	-0.223	0.224
Colon	38	HM450k	-0.303	0.304
Kidney	205	HM450k	-0.359	0.359
Ovary	12	HM27k	-0.193	0.19
<i>RMSE</i> root mean square error; <i>HM450k</i> Illumina HumanMethylation450 BeadChip; <i>HM27k</i> Illumina HumanMethylation27 BeadChip				

To further assess model performance while mitigating platform-related bias, we performed additional external validation using CPTAC adjacent normal samples generated on the EPIC array, consistent with the array used for model training. Due to limited tissue availability, only kidney and lung samples were included. Validation results are summarized in Table 3. The results revealed minimal bias (-0.009 for lung and -0.028 for kidney) and very low RMSE values (0.021 for lung and 0.029 for kidney), indicating that the predicted tissue-specific hypomethylation score closely matched observed global PMD hypomethylation.

Table 3
External validation of tissue-specific mitotic clocks on CPTAC adjacent normal samples

Tissue	Number of Samples	Array	Performance metrics	
			Bias	RMSE
Lung	280	EPIC	-0.009	0.021
Kidney	159	EPIC	-0.028	0.029
<i>RMSE</i> root mean square error; <i>EPIC</i> Illumina Infinium MethylationEPIC BeadChip				

Tissue-specific mitotic clocks correlate with lifetime stem cell divisions of a given tissue and cancer risk

Tomasetti and Vogelstein previously reported that lifetime cancer risk across tissues is strongly associated with TNSD occurring over a tissue's lifetime (*Iscd*)^{1,2}. *Iscd* reflects both the total number of stem cells in a fully developed tissue, and TNSD for each stem cell undergoes over a lifetime. Because direct quantification of *Iscd* in vivo is challenging, they employed a mathematical approach based on experimentally observed data to derive estimates, including tracing cell lineage with markers (like viral tags), and functional assays such as the mammosphere assay, which assess self-renewal rates in vitro, and linked observed cell numbers and division frequencies to total lifetime activity and cancer risk^{1,2}. Reported *Iscd* values for selected tissues were: breast (*Iscd* = 3.03×10^{12}), ovary (*Iscd* = 2.20×10^7), prostate (*Iscd* = 5.08×10^{10}), lung (*Iscd* = 9.27×10^9), and colon (*Iscd* = 1.17×10^{12}). To determine whether our tissue-specific mitotic clock and other established mitotic clocks can capture this tissue-level mitotic burden, we applied each clock to multiple TCGA adjacent normal samples per tissue and averaged the estimates to derive tissue-level proxies of *Iscd*. As shown in Fig. 2, the tissue-specific mitotic clocks exhibited a strong association with experimentally derived *Iscd* (*cor* = -0.745, *P*-value = 0.15), whereas other existing mitotic clocks, including epiTOC1, epiTOC2, and stemTOC, showed weak correlations (*cor* = 0.14, 0.01, 0.2 respectively). These results indicate that the tissue-specific clocks can effectively reflect TNSD at the tissue level. Notably, the observed trends are biologically coherent and markedly stronger than those produced by existing comparator clocks.

We further examined whether the estimated *Iscd* was associated with cancer risk to test the hypothesis by Tomasetti and Vogelstein^{1,2}. As shown in Fig. 3, the tissue-specific mitotic clock estimates demonstrated a moderate positive association with cancer incidence (*cor* = 0.32, *P*-value = 0.53), whereas other established mitotic clocks showed no meaningful correlations (*cor* = 0.09, -0.15, 0.06 respectively). Although this association did not reach statistical significance, the observed trend supports the hypothesis that cumulative mitotic activity contributes to tissue-specific cancer risk and further highlights the biological relevance of the tissue-specific mitotic clock. In contrast, the lack of association observed for the other mitotic clocks suggests that they may be less sensitive to tissue-specific variation in cumulative mitotic activity relevant to cancer risk.

Discussion

In this study, we developed and validated tissue-specific mitotic clocks to estimate TNSD for six human tissues. By leveraging global hypomethylation at PMD solo-WCGW sites as a biological surrogate of mitotic activity – rather than relying on chronological age – our clocks significantly outperformed established mitotic clocks in capturing the distinct proliferative histories of different organs. Specifically, our models exhibited minimal estimation bias when validated in an independent, array-consistent CPTAC cohort.

Our tissue-specific mitotic clocks generated directionally consistent associations with reported *lscd* and tissue-specific cancer incidence, supported by strong effect sizes and rank-order concordance. Although these relationships were statistically underpowered due to the limited number of tissue types available for analysis. In this context, the biological coherence and effect magnitude, rather than statistical significance, are most informative, and both were markedly stronger for our tissue-specific mitotic clocks than for existing comparator clocks. Chronological age-trained clocks primarily capture systemic aging signals. In contrast, our approach targets replication-coupled epigenetic erosion and loss of epigenetic integrity. This biological distinction likely explains why existing mitotic clocks perform well for age prediction but poorly for correlating tissue-level cancer risk.

The improved performance supports our central assumption. Unlike existing mitotic clocks, which are largely pan-tissue and assume a universal set of CpG loci behaves consistently across tissues, our study based on the premise that different tissues exhibit unique intrinsic aging rates and molecular signatures. As a result, a universal model insufficient to capture this biological heterogeneity. We empirically support this hypothesis through PCA of 836 tissue samples from 421 GTEx donors (Fig. 1). PMD methylation profiles robustly distinguished tissue types, suggesting that inter-tissue epigenetic differences reflect variation in biological aging dynamics rather than stochastic noise. To account for the potential confounding of chronological age, we further analyzed matched breast, colon, and ovary tissues from 10 donors (Figure S1). The persistence of clear tissue separation in this controlled, intra-individual setting confirms that these epigenetic differences are tissue-intrinsic rather than age-dependent. These observations align with prior findings by Pierce et al.^{22,23} and Schaefer et al.²⁴, which also highlighted significant tissue-specific heterogeneity in DNA methylation aging signatures. Collectively, our results demonstrate that accurate quantification of proliferative history requires tissue-specific mitotic clocks rather than a global, pan-tissue model.

Notably, despite the clear tissue-specific clustering observed in the PCA of matched tissue samples (Figure S1), global PMD hypomethylation levels were not statistically significantly different across the same cohort (Figure S2). Although this lack of significance in global comparisons may partly reflect limited statistical power due to the small sample size ($N = 10$), it more fundamentally highlights a methodological limitation: averaging thousands of CpG measurements into a single global metric could dilute biological signal and reduce sensitivity to biologically meaningful differences. Conversely, multi-dimensional reduction techniques, such as PCA, retain the covariance structure across all CpG sites, allowing for the detection of subtle yet consistent patterns that are often masked in univariate analyses, particularly in studies with limited sample sizes. This interpretation is supported by prior studies. Farré et

al. demonstrated that PCA can identify major sources of variation (like tissue type, cell type, age) across methylation profiles on the matched brain and blood samples from 17 individuals²⁹. Furthermore, Eulalio et al. developed *regionalpcs* approach, which uses PCA to summarize methylation across CpG clusters and achieved a ~ 54% improvement in sensitivity over mean-based metrics³⁰. Together, these findings suggest that reliance on global methylation averages may underestimate tissue-specific epigenetic heterogeneity, reinforcing the advantage of multi-dimensional approach for detecting biologically meaningful differences.

During external validation of our tissue-specific mitotic clocks on TCGA adjacent normal data, we observed a consistent systematic bias across tissues. This bias might be attributed to two primary factors. First, the tissue-specific mitotic clocks capture tissue-level methylation variability that is not fully represented by the global PMD hypomethylation, leading to deviations from the global PMD estimate. Second, TCGA methylation data were primarily generated using the HM450k array, which has limited probe coverage compared to the EPIC array used for model training, thereby introducing systematic bias due to missing CpG measurements. To mitigate this technical effect and more accurately assess model performance, we further validated our models on additional CPTAC adjacent normal samples profiled on the EPIC array. In this platform-consistent setting, systematic bias was largely shrunk, demonstrating that array consistency substantially improves model prediction accuracy. These results confirm that our tissue-specific mitotic clocks are highly robust and generalizable on independent datasets generated with compatible platforms. Accordingly, the bias observed in TCGA validation is most plausibly attributed to limitations of the legacy HM450k design, which lacks the probe density required for precise quantification, although such datasets remain informative for capturing broad directional trends. Array technologies are converging asymptotically toward a stable measurement regime, with marginal performance gains across successive updates. As next-generation, higher-density arrays such as EPIC continue to be widely adopted, platform-related bias will become negligible, ensuring the robustness and broad applicability of our mitotic clock models. While whole-genome bisulfite sequencing provides a useful technical reference for methylation measurement, the degree to which array-based platforms approximate sequencing-level resolution remains an open area for future systematic benchmarking.

Our study has several notable strengths. First, it represents the first systematic effort to construct tissue-specific DNA methylation-based mitotic clocks using biologically “normal” samples from the GTEx project. Unlike prior mitotic clocks derived from TCGA adjacent normal tissues, the precision of our training data minimizes potential confounding from field effects^{25–27} or pre-neoplastic epigenetic alterations²⁸. Field cancerization refers to the presence of molecularly altered but histologically normal cells surrounding tumors, arising from early clonal expansions during tumorigenesis³¹. These altered fields can exhibit elevated proliferative activity and replication-associated epigenetic and genetic changes³². As a result, adjacent “normal” tissues may carry inflated mitotic signals that do not reflect baseline tissue turnover. Training mitotic clocks on TCGA normal-adjacent samples may therefore bias TNSD estimates upward and distort tissue-specific proliferative histories.

Second, by modeling global PMD hypomethylation at solo-WCGW CpGs, our approach captures a biologically validated and experimentally supported biomarker of TNSD, rather than relying on chronological age as a surrogate. Finally, our tissue-specific estimates exhibit stronger associations with reported lscd and cancer risk compared to existing clocks, providing empirical support for the Tomasetti and Vogelstein hypothesis linking cumulative stem cell divisions to cancer susceptibility.

Nevertheless, several limitations should be acknowledged. Sample sizes for breast and kidney training sets were modest, which may constrain model precision. Additionally, while our colon model was trained exclusively on transverse colon samples, the TCGA validation cohort included a mixture of ascending, transverse, and descending colon tissues, which could contribute to the observed bias. Lastly, although correlation analyses revealed strong associations with reported lscd and cancer risk, the evaluation was limited to six tissue types, reducing the statistical power to fully characterize the relationship between TNSD and cancer incidence across the human body.

Beyond methodological advances, tissue-specific mitotic clocks also have several potential applications. They may help identify tissues or individuals with unusually high proliferative burden, indicating elevated baseline cancer susceptibility and supporting risk stratification. In addition, these measurements may help inform cancer prevention strategies by highlighting organs in which reducing proliferative stress or enhancing genomic maintenance could yield the greatest benefit. In addition, our mitotic clocks provide a quantitative baseline of replication-driven risk, enabling separation of intrinsic proliferative contributions from external exposure effects. By establishing the expected cancer risk attributable to endogenous cell turnover, this framework supports more precise estimation of the incremental risk associated with environmental or behavioral factor. Together, these applications position tissue-specific mitotic clocks as a tool for both mechanistic insight and population-level risk assessment.

In summary, we demonstrate that tissue-specific mitotic clocks provide a more accurate measure of proliferative history than pan-tissue models. This finding highlights the importance of accounting for tissue heterogeneity in modeling mitotic aging. Future work could expand this framework to larger cohorts and more diverse tissue panels will be critical for improving model generalizability and biological relevance, as well as for more precisely characterizing the relationship between TNSD and cancer risk.

Methods

GTEx data for model development

GTEx project is a large-scale, publicly accessible biobank designed to investigate the association between genetic variation and gene expression across human tissues³³. It includes samples from approximately 960 postmortem donors who predominantly died from acute, non-disease-related causes³³. Each donor contributed multiple tissue types and only samples confirmed to be free of significant disease determined by pathological examination and medical record review were included to ensure the representation of true normal tissues³⁴. The GTEx v8 release comprises 15,201 RNA-seq

samples from 838 donors across 49 tissue types, with detailed demographic and technical metadata³⁴. The enhancing GTEx (eGTEx) project expands this resource by adding epigenomic and proteomic profiling. Among these, DNA methylation was measured using the Illumina Infinium MethylationEPIC BeadChip (EPIC)³⁵.

We obtained eGTEx methylation data from the Gene Expression Omnibus (GEO; accession number GSE213478³⁶). Quality control and preprocessing were performed by Pierce *et al.*²² The dataset contains 754,119 probes across 987 samples from 9 normal tissue types. For model construction, we selected six tissues representing common sites of cancer origin: breast mammary tissue (n = 52), colon transverse (n = 224), kidney cortex (n = 50), lung (n = 223), ovary (n = 164), and prostate (n = 123). All samples were derived from 421 donors in the GTEx cohort.

TCGA and CPTAC data for external validation

We downloaded DNA methylation data of adjacent normal tissue for six selected tissue types from TCGA via the Genomic Data Commons (GDC) portal³⁷. Although TCGA provides preprocessed data, additional quality control was performed to exclude low-quality samples and probes with detection *p*-value > 0.05. After quality control, the external validation cohort included female breast (n = 97), colon (n = 38), kidney (n = 205), lung (n = 74), ovary (n = 12) and prostate (n = 50). All samples were profiled using the Illumina HumanMethylation450 BeadChip (HM450k), except for ovarian tissue, which was generated using the Illumina HumanMethylation27 (HM27k) platform.

To ensure consistency with our Illumina Infinium EPIC array-based training model, we further obtained an additional external validation dataset from the Clinical Proteomic Tumor Analysis Consortium (CPTAC)³⁸. CPTAC extends TCGA by performing deep proteomic characterization of tumor and adjacent normal samples, complementing genomic and epigenomic data to enable integrated proteogenomic analyses of cancer³⁸. Because CPTAC profiled a limited number of tissues, only kidney (n = 280) and lung (n = 159) adjacent normal samples were included. In both validation datasets, DNA methylation levels were represented as β -values.

Calculation of global PMD hypomethylation

Global DNA hypomethylation was quantified using PMD solo-WCGW CpG sites. Although PMD solo-WCGWs were initially identified using WGBS, Laird *et al.* demonstrated that methylation levels at these sites measured on microarray platforms (e.g. HM450k) strongly correlate with WGBS measurements and reliably capture PMD hypomethylation patterns despite limited probe coverage²⁰. Building on this work, we used a curated set of 26,732 PMD solo-WCGW CpGs present on EPIC array. Specifically, only probes targeting solo-WCGW CpGs were retained, while those located within annotated CpG islands or exhibiting low methylation levels ($\beta < 0.2$) were excluded²⁰. For each sample, the average β -value across the retained probes was computed to derive global PMD hypomethylation, reflecting large-scale

methylation loss across PMDs. This measure was used as the outcome variable in our model development.

Tissue-specific mitotic clock construction

We constructed tissue-specific mitotic clocks using the following strategy: for each tissue type, samples were randomly partitioned into training (70%) and testing (30%) subsets. An elastic net regression model with alpha parameter = 0.5 (R package 'glmnet') was trained to predict the global PMD hypomethylation, derived from CpG β -values within each tissue. Sex was included as an unpenalized covariate to account for sex-related effects, except in inherently sex-specific tissues such as prostate and ovary. For breast tissue, only female samples were analyzed given the markedly lower incidence rate of breast cancer in males. Model robustness was evaluated using five-fold cross-validation. The final set of CpGs selected for each tissue type was used to construct tissue-specific mitotic clocks. Coefficients for each tissue-specific model are provided in Supplementary Table 1.

Statistical analysis

All statistical analyses were performed using R version 4.3.2. Tissue-specific methylation variation was evaluated using principal component analysis (PCA) on PMD solo-WCGWs beta values. Global PMD hypomethylation differences across matched tissues were assessed using repeated-measures analysis of variance (RMANOVA).

Missing values in CPTAC methylation data were imputed with impute.knn (R Bioconductor package 'impute')³⁹. For external validation, the global PMD hypomethylation served as the observed outcome, and predicted tissue-specific hypomethylation score as predicted values; model performance was quantified using estimation bias and root mean squared error (RMSE). Pearson correlation was used to evaluate associations between mitotic clock estimates and experimentally derived lifetime stem cell divisions (lscd), as well as cancer incidence rates. Age-adjusted cancer incidence rates were obtained from the Surveillance, Epidemiology, and End Results (SEER) database⁴⁰. Existing mitotic clocks, including epiTOC1, epiTOC2, and stemTOC2, were computed using the EpiMitClocks R package¹⁴. Statistical significance was defined as P-value < 0.05.

Declarations

Competing interests

The authors declare that they have no competing interests.

Author contributions

F.R.S. conceptualized the study. Y.W. developed method development and all statistical analyses. All authors contributed to the interpretation of the findings. Y.W. and F.R.S. both contributed to the writing, reviewing, and editing.

Materials & Correspondence

Correspondence and requests for materials should be addressed to Fredrick R. Schumacher

References

1. Tomasetti C, Vogelstein B (2015) Variation in cancer risk among tissues can be explained by the number of stem cell divisions. *Science* 347:78–81
2. Tomasetti C, Li L, Vogelstein B (2017) Stem cell divisions, somatic mutations, cancer etiology, and cancer prevention. *Science* 355:1330–1334
3. Tomasetti C et al (2017) Role of stem-cell divisions in cancer risk. *Nature* 548:E13–E14
4. Minter CJ et al (2023) More than bad luck: Cancer and aging are linked to replication-driven changes to the epigenome. *Sci Adv* 9:eadf4163
5. Giovannucci EL (2016) Are Most Cancers Caused by Specific Risk Factors Acting on Tissues With High Underlying Stem Cell Divisions? *J Natl Cancer Inst* 108:djv343
6. Hansen KD et al (2011) Increased methylation variation in epigenetic domains across cancer types. *Nat Genet* 43:768–775
7. Du Q et al (2021) DNA methylation is required to maintain both DNA replication timing precision and 3D genome organization integrity. *Cell Rep* 36
8. Alexandrov LB et al (2015) Clock-like mutational processes in human somatic cells. *Nat Genet* 47:1402–1407
9. Kim JY, Tavaré S, Shibata D Counting human somatic cell replications: Methylation mirrors endometrial stem cell divisions. *Med Sci*
10. Duran-Ferrer M et al (2020) The proliferative history shapes the DNA methylome of B-cell tumors and predicts clinical outcome. *Nat Cancer* 1:1066–1081
11. Youn A, Wang S (2018) The MiAge Calculator: a DNA methylation-based mitotic age calculator of human tissue types. *Epigenetics* 13:192–206
12. Yang Z et al (2016) Correlation of an epigenetic mitotic clock with cancer risk. *Genome Biol* 17:205
13. Teschendorff AE (2020) A comparison of epigenetic mitotic-like clocks for cancer risk prediction. *Genome Med* 12:56
14. Zhu T, Tong H, Du Z, Beck S, Teschendorff AE (2024) An improved epigenetic counter to track mitotic age in normal and precancerous tissues. *Nat Commun* 15:4211
15. López-Otín C, Blasco MA, Partridge L, Serrano M, Kroemer G (2013) The hallmarks of aging. *Cell* 153:1194–1217
16. López-Otín C, Blasco MA, Partridge L, Serrano M, Kroemer G (2023) Hallmarks of aging: An expanding universe. *Cell* 186:243–278
17. Bogeska R et al (2022) Inflammatory exposure drives long-lived impairment of hematopoietic stem cell self-renewal activity and accelerated aging. *Cell Stem Cell* 29:1273–1284e8

18. Xu C, Luo J, He L, Montell C, Perrimon N (2017) Oxidative stress induces stem cell proliferation via TRPA1/RyR-mediated Ca²⁺ + signaling in the *Drosophila* midgut. *eLife* 6, e22441
19. Scodellaro C et al (2024) Unlocking the Potential of Stem Cell Microenvironments In Vitro. *Bioengineering* 11
20. Zhou W et al (2018) DNA methylation loss in late-replicating domains is linked to mitotic cell division. *Nat Genet* 50:591–602
21. Endicott JL, Nolte PA, Shen H, Laird PW (2022) Cell division drives DNA methylation loss in late-replicating domains in primary human cells. *Nat Commun* 13:6659
22. Jain N et al (2024) DNA methylation correlates of chronological age in diverse human tissue types. *Epigenetics Chromatin* 17:25
23. Richardson M et al (2025) Characterization of DNA methylation clock algorithms applied to diverse tissue types. *Aging* 17:67–96
24. Dmitrijeva M, Ossowski S, Serrano L, Schaefer MH (2018) Tissue-specific DNA methylation loss during ageing and carcinogenesis is linked to chromosome structure, replication timing and cell division rates. *Nucleic Acids Res* 46:7022–7039
25. Yates J et al (2024) DNA-methylation variability in normal mucosa: a field cancerization marker in patients with adenomatous polyps. *J Natl Cancer Inst* 116:974–982
26. Teschendorff AE et al (2016) DNA methylation outliers in normal breast tissue identify field defects that are enriched in cancer. *Nat Commun* 7:10478
27. Møller M et al (2017) Heterogeneous patterns of DNA methylation-based field effects in histologically normal prostate tissue from cancer patients. *Sci Rep* 7:40636
28. Saavedra KP, Brebi PM, Roa JC (2012) Epigenetic alterations in preneoplastic and neoplastic lesions of the cervix. *Clin Epigenetics* 4:13
29. Farré P et al (2015) Concordant and discordant DNA methylation signatures of aging in human blood and brain. *Epigenetics Chromatin* 8:19
30. Eulalio T et al (2025) regionalpcs improve discovery of DNA methylation associations with complex traits. *Nat Commun* 16:368
31. Willenbrink TJ et al (2020) Field cancerization: Definition, epidemiology, risk factors, and outcomes. *J Am Acad Dermatol* 83:709–717
32. Demaria S et al (2010) Cancer and inflammation: promise for biologic therapy. *J Immunother* 33:335–351
33. Lonsdale J et al (2013) The Genotype-Tissue Expression (GTEx) project. *Nat Genet* 45:580–585
34. The GTEx Consortium atlas of genetic regulatory effects across human tissues (2020)
35. Stranger BE et al (2017) Enhancing GTEx by bridging the gaps between genotype, gene expression, and disease. *Nat Genet* 49:1664–1670
36. Oliva M et al (2023) DNA methylation QTL mapping across diverse human tissues provides molecular links between genetic variation and complex traits. *Nat Genet* 55:112–122

37. Weinstein JN et al (2013) The Cancer Genome Atlas Pan-Cancer analysis project. *Nat Genet* 45:1113–1120

38. Rudnick PA et al (2016) A Description of the Clinical Proteomic Tumor Analysis Consortium (CPTAC) Common Data Analysis Pipeline. *J Proteome Res* 15:1023–1032

39. Troyanskaya O et al (2001) Missing value estimation methods for DNA microarrays. *Bioinformatics* 17:520–525

40. Surveillance Epidemiology, and End Results Program. *SEER* <https://seer.cancer.gov/index.html>

Table 1

Table 1 is available in the Supplementary Files section.

Figures

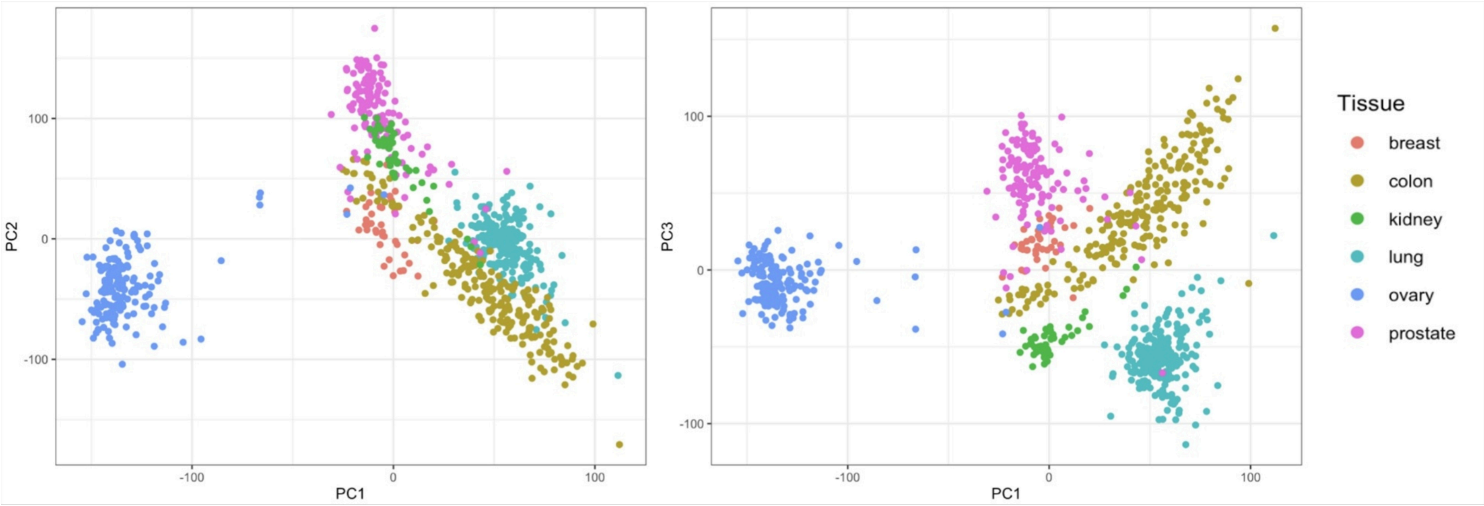


Figure 1

Principal component analysis of PMD solo-WCGW CpGs across six tissues

Note: PC1 represents the first principal component, PC2 represents the second principal component, PC3 represents the third principal component

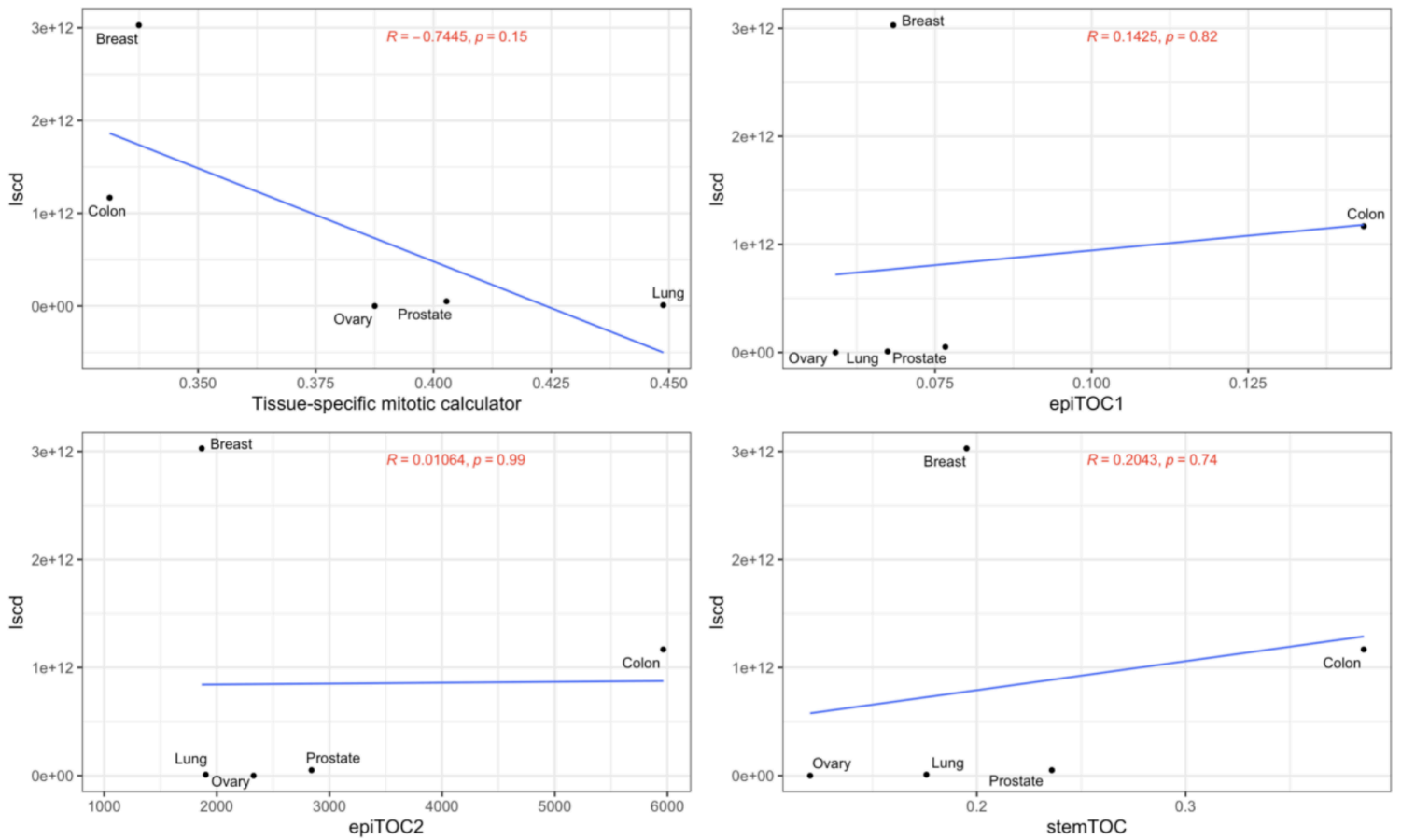


Figure 2

Correlation between mitotic clock estimates and experimentally derived lifetime stem cell divisions across five tissues

Note: lscd represents the total number of stem cell divisions occurring over a tissue's lifetime

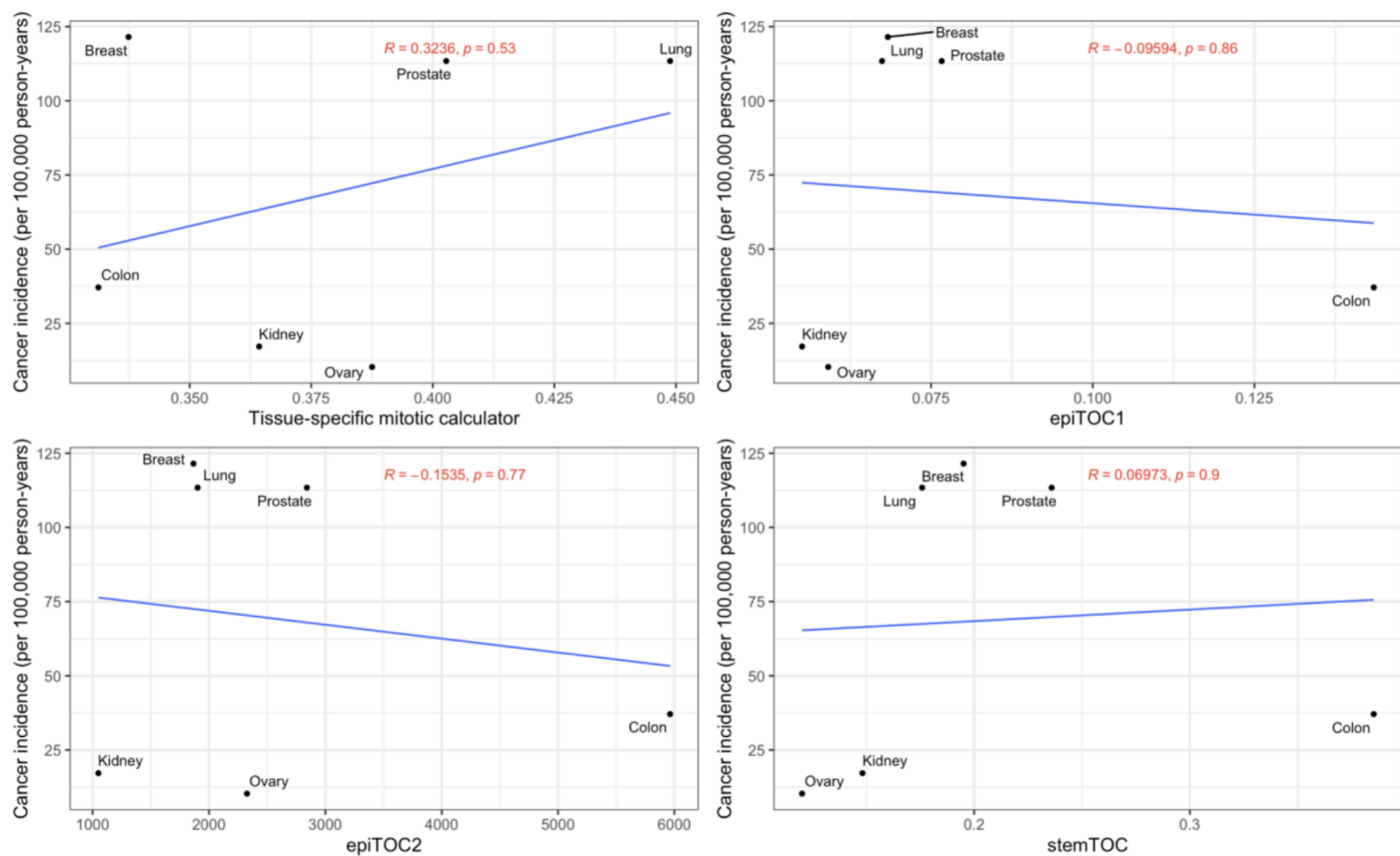


Figure 3

Correlation between mitotic clock estimates and cancer incidence

Supplementary Files

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

- [SupplementaryTable1.xlsx](#)
- [SupplementaryTable2.csv](#)
- [Supplementaryinformation.pdf](#)
- [Table1.docx](#)