Title

Sperm epigenetic clock associates with pregnancy outcomes in the general population.

Journal

Human Reproduction, 37(7)

ISSN

0268-1161

Authors

Pilsner, J Richard
Saddiki, Hachem
Whitcomb, Brian W
et al.

Publication Date

2022-06-30

DOI

10.1093/humrep/deac084

Peer reviewed

Human Reproduction, Vol.37, No.7, pp. 1581–1593, 2022
Advance Access Publication on May 13, 2022 https://doi.org/10.1093/humrep/deac084

human reproduction

ORIGINAL ARTICLE *Reproductive epidemiology*

Sperm epigenetic clock associates with pregnancy outcomes in the general population

J. Richard Pilsner [1,2,*], Hachem Saddiki[3], Brian W. Whitcomb [3], Alexander Suvorov[4], Germaine M. Buck Louis[5], Sunni L. Mumford [6], Enrique F. Schisterman[6], Oladele A. Oluwayiose[1], and Laura B. Balzer [3,*]

[1]Department of Obstetrics and Gynecology, Wayne State University, Detroit, MI, USA [2]Institute of Environmental Health Sciences, Wayne State University, Detroit, MI, USA [3]Department of Biostatistics and Epidemiology, School of Public Health and Health Sciences, University of Massachusetts Amherst, Amherst, MA, USA [4]Environmental Health Sciences, School of Public Health and Health Sciences, University of Massachusetts Amherst, Amherst, MA, USA [5]College of Health and Human Services, George Mason University, Fairfax, VA, USA [6]Department of Biostatistics, Epidemiology and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

*Correspondence address. Department of Obstetrics and Gynecology, Wayne State University, 275 East Hancock, Detroit, MI 48201, USA. E-mail: rpilsner@wayne.edu https://orcid.org/0000-0002-4541-6267 (J.R.P.); Department of Biostatistics and Epidemiology, School of Public Health and Health Sciences, University of Massachusetts Amherst, 715 N. Pleasant Street, Amherst, MA 01003, USA. E-mail: lbalzer@umass.edu https://orcid.org/000-0002-3730-410X (L.B.B.)

Submitted on December 14, 2021; resubmitted on March 28, 2022; editorial decision on April 7, 2022

STUDY QUESTION: Is sperm epigenetic aging (SEA) associated with probability of pregnancy among couples in the general population?

SUMMARY ANSWER: We observed a 17% lower cumulative probability at 12 months for couples with male partners in the older compared to the younger SEA categories.

WHAT IS KNOWN ALREADY: The strong relation between chronological age and DNA methylation profiles has enabled the estimation of biological age as epigenetic 'clock' metrics in most somatic tissue. Such clocks in male germ cells are less developed and lack clinical relevance in terms of their utility to predict reproductive outcomes.

STUDY DESIGN, SIZE, DURATION: This was a population-based prospective cohort study of couples discontinuing contraception to become pregnant recruited from 16 US counties from 2005 to 2009 and followed for up to 12 months.

PARTICIPANTS/MATERIALS, SETTING, METHODS: Sperm DNA methylation from 379 semen samples was assessed via a beadchip array. A state-of-the-art ensemble machine learning algorithm was employed to predict age from the sperm DNA methylation data. SEA was estimated from clocks derived from individual CpGs (SEA_{CpG}) and differentially methylated regions (SEA_{DMR}). Probability of pregnancy within 1 year was compared by SEA, and discrete-time proportional hazards models were used to evaluate the relations with time-to-pregnancy (TTP) with adjustment for covariates.

MAIN RESULTS AND THE ROLE OF CHANCE: Our SEA_{CpG} clock had the highest predictive performance with correlation between chronological and predicted age ($r = 0.91$). In adjusted discrete Cox models, SEA_{CpG} was negatively associated with TTP (fecundability odds ratios (FORs)=0.83; 95% CI: 0.76, 0.90; $P = 1.2 \times 10^{-5}$), indicating a longer TTP with advanced SEA_{CpG}. For subsequent birth outcomes, advanced SEA_{CpG} was associated with shorter gestational age (n = 192; −2.13 days; 95% CI: −3.67, −0.59; $P = 0.007$). Current smokers also displayed advanced SEA_{CpG} ($P < 0.05$). Finally, SEA_{CpG} showed a strong performance in an independent IVF cohort (n = 173; $r = 0.83$). SEA_{DMR} performance was comparable to SEA_{CpG} but with attenuated effect sizes.

LIMITATIONS, REASONS FOR CAUTION: This prospective cohort study consisted primarily of Caucasian men and women, and thus analysis of large diverse cohorts is necessary to confirm the associations between SEA and couple pregnancy success in other races/ethnicities.

WIDER IMPLICATIONS OF THE FINDINGS: These data suggest that our sperm epigenetic clocks may have utility as a novel biomarker to predict TTP among couples in the general population and underscore the importance of the male partner for reproductive success.

STUDY FUNDING/COMPETING INTEREST(S): This work was funded in part by grants the National Institute of Environmental Health Sciences, National Institutes of Health (R01 ES028298; PI: J.R.P. and P30 ES020957); Robert J. Sokol, MD Endowed Chair of Molecular Obstetrics and Gynecology (J.R.P.); and the Intramural Research Program of the Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, Maryland (Contracts N01-HD-3-3355, N01-HD-3-3356 and N01-HD-3-3358). S.L.M. was supported by the Intramural Research Program of the Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health. The authors declare no competing interests.

TRIAL REGISTRATION NUMBER: N/A.

Key words: sperm epigenetic clock / aging / methylation / infertility / pregnancy / machine learning

Introduction

Infertility, a disease of the reproductive system defined by the failure to achieve a clinical pregnancy after 12 months or more of regular unprotected sexual intercourse (Cooper *et al.*, 2010), is estimated to affect as many as 17% of couples seeking to have children in the USA and other developed countries worldwide (Boivin *et al.*, 2007; Smith *et al.*, 2011; Thoma *et al.*, 2013). While infertility has been primarily treated as a female predicament, around one-half of infertility cases can be tracked to male factors (Whitman-Elia and Baxley, 2001). Clinically, male infertility is typically determined using measures of semen quality recommended by World Health Organization (WHO) cut points (Cooper *et al.*, 2010). A major limitation, however, is that standard semen analyses are relatively poor predictors of reproductive capacity and success (Jungwirth et al., 2012; Buck Louis *et al.*, 2014). Despite major advances in understanding the molecular and cellular functions in sperm over the last several decades, semen analyses remain the primary method to assess male fecundity and fertility.

Chronological age is a significant determinant of human fecundity and fertility. The disease burden of infertility is likely to continue to rise as parental age at the time of conception has been steadily increasing (Waldenström, 2016). While the emphasis has been on the effects of advanced maternal age on adverse reproductive and offspring health (Hemminki and Kyyrönen, 1999; Jacobsson *et al.*, 2004; Kemkes-Grottenthaler, 2004), new evidence suggests that, irrespective of maternal age, higher male age contributes to longer time-to-conception (Hassan and Killick, 2003), poor pregnancy outcomes (Hassan and Killick, 2003; Horta *et al.*, 2019; Oluwayiose *et al.*, 2021) and adverse health of the offspring in later life (Montgomery *et al.*, 2004; Saha *et al.*, 2009; Puleo *et al.*, 2012). However, chronological age does not capture the internal (e.g. genetics) and external factors (e.g. environmental conditions) that may affect cellular aging processes. Therefore, chronological age is a proxy measure of the actual biological age of cells.

The effect of chronological age on the genomic landscape of DNA methylation is profound (Jenkins *et al.*, 2018; Oluwayiose *et al.*, 2021) and likely occurs through the accumulation of maintenance errors of DNA methylation over the lifespan, which have been originally described as epigenetic drift (Fraga *et al.*, 2005). In recent years, the strong relation between age and DNA methylation profiles has enabled the development of statistical models to estimate biological age in most somatic tissue via different epigenetic 'clock' metrics, such as DNA methylation age and epigenetic age acceleration, which describe the degree to which predicted biological age deviates from chronological age (Hannum *et al.*, 2013; Horvath, 2013; Levine *et al.*, 2018). In turn, these epigenetic clock metrics have emerged as novel biomarkers of a host of phenotypes such as allergy and asthma in children (Peng *et al.*, 2019), early menopause (Levine *et al.*, 2016), increased incidence of cancer types and cardiovascular-related diseases (Fransquet *et al.*, 2019), frailty (Breitling *et al.*, 2016) and cognitive decline in adults (Degerman *et al.*, 2017). They also display good predictive ability for cancer, cardiovascular and all-cause mortality (Perna *et al.*, 2016).

While epigenetic clock metrics are powerful tools to better understand the aging process in somatic tissue as well as their associations with adverse disease outcomes and mortality, the DNA methylation loci used for these metrics have shown no predictive value in male germ cells (Horvath, 2013). To date, only a few studies have constructed epigenetic clocks specific to male germ cells (Jenkins *et al.*, 2018; Cao *et al.*, 2020; Laurentino *et al.*, 2020) and only one study by Jenkins *et al.* (2018) reported that smokers trended toward an increased epigenetic age compared to non-smokers. These results indicate that sperm epigenetic clocks hold promise as a novel biomarker for reproductive health and/or environmental exposures. However, the relation between sperm epigenetic clocks and reproductive outcomes has not been examined. Consequently, this study utilized sperm EPIC array data to construct a novel epigenetic clock via an ensemble machine learning method to predict chronological age from 379 male partners of couples recruited from the general population who were discontinuing contraception for purposes of becoming pregnant. We then determined the associations of sperm epigenetic age (SEA) with couples' reproductive outcomes, including time-to-pregnancy (TTP) and birth outcomes, and, as a secondary aim, with male smoking.

Materials and methods

Study populations

Banked whole semen samples were obtained from male participants of the Longitudinal Investigation of Fertility and the Environment (LIFE) Study, a prospective pregnancy cohort for which details have been previously published (Buck Louis *et al.*, 2011; Buck Louis *et al.*, 2014). Briefly, 501 couples were recruited from 16 counties in Michigan and Texas, USA, using a population-based sampling frame (Buck Louis *et al.*, 2011). Eligibility criteria were: in a committed relationship and planning to discontinue contraception to become pregnant; females aged 18–40 years and males aged 18 years or older; no injectable contraceptive use in the past year or off contraception for >2 months; females' self-reported menstrual cycles between 21–42 days; and an ability to communicate in English or Spanish (Buck Louis *et al.*, 2014). Couples with physician-diagnosed infertility were ineligible for enrollment. The current study includes 379 (76%) couples who had a

remaining aliquot of semen available for DNA methylation analyses. Study participants gave written informed consent before any data collection, and full institutional review board approval for human subjects was received from all collaborating institutions.

To assess the generalizability of our clocks through their application to predict chronological age from sperm methylation, we utilized existing data (n = 173) from the Sperm Environmental Epigenetics and Development Study (SEEDS), a prospective observational cohort study aimed at investigating the associations of male preconception endocrine disrupting chemical exposure with sperm epigenetics and subsequent early-life development (Wu *et al.*, 2017; Oluwayiose *et al.*, 2021). Participants were recruited from couples undergoing fertility treatment at Baystate Medical Center located in Springfield, MA, USA. The inclusion criteria were male partners 18–55 years old without vasectomy and fresh ejaculate sperm used for IVF treatment.

Data collection

Semen, blood and urine biospecimens were collected in the homes of each participant, as previously described (Buck Louis *et al.*, 2014). For men, whole semen samples were collected at entry into the study and a second sample 1 month later. The first semen samples were utilized for this current analysis. Both samples were collected, after a minimal 2-day period of abstinence, via masturbation without the use of any lubricant. Details regarding the sample collection procedures, shipping materials provided to participants, and semen parameter quantification methods using Computer Assisted Semen Analysis have been described in detail in prior analyses of the LIFE Study (Buck Louis *et al.*, 2014; Bloom *et al.*, 2015).

Women were instructed to use the Clearblue Easy^{TM} Fertility Monitor to time intercourse relative to ovulation in order to maximize timed intercourse. Women were followed daily up to 12 months of trying. TTP was considered as the number of menstrual cycles until the hCG result confirmed pregnancy (Buck Louis *et al.*, 2011). Women not becoming pregnant were censored at 12 months. Pregnant women were followed until delivery when they completed and returned birth announcements that captured infant's date of birth, sex, birthweight (in grams), birth length (in centimeters) and head circumference (in centimeters). We defined gestational age (GA) as the interval from the date of estimated conception, which we assumed was the day of ovulation as indicated by the peak LH reading on the fertility monitor to date of delivery. As such, we have post-conception GA, which is approximately 2 weeks shorter than GA based on last menstrual period (LMP). Thus, our preterm delivery outcome is defined as GA <245 days (35 weeks) from the date of ovulation (conventionally defined as 37 weeks from LMP date in the absence of ovulation data (1977)), which is an accurate proxy for timing of conception related to the limited period of viability of the oocyte (Wilcox *et al.*, 1995). We calculate the infant's ponderal index, an indicator of anthropometric proportionality (Landmann *et al.*, 2006), using the formula: birth length/birthweight^3 × 100.

Sample preparation and DNA isolation

To separate sperm from seminal plasma and somatic cells, semen samples underwent a one-step 40% gradient centrifugation. Sperm DNA was isolated with our previously published method (Wu *et al.*, 2015), which homogenizes sperm in the presence of 0.2 mm steel beads, RLT buffer (Qiagen, Hilden, Germany), and 50 mM of tris(2-carboxyethyl)phosphine (TCEP; Pierce, Rockford, IL, USA) prior to the isolation of sperm DNA via silica-column purification.

Identification of age-associated sperm methylation

Sperm DNA methylation was assessed using EPIC Infinium Methylation Beadchip (Illumina, San Diego, CA, USA), which allows for genome-wide coverage of over 850 000 methylation sites. Batch effects were minimized by randomizing samples within and across the beadchip. The beadchip was run at the Yale Center for Genome Analyses at Yale University. Within-array normalization and dye-bias equalization of Type I and Type II probes (Niu *et al.*, 2016) was performed using the normal-exponential convolution method (Noob) (Triche *et al.*, 2013) via minfi package in R (Aryee *et al.*, 2014). Batch effects were corrected with the ComBat function in the sva package (Leek *et al.*, 2012) while cross-hybridizing probes were removed using the DMRcate package (Peters *et al.*, 2015). Of the total 865 859 individual CpG sites interrogated for each participant, 62 796 sites were excluded after preprocessing, leaving a total of 803 063 CpG sites available for downstream analyses.

To verify our methylation data were not influenced by somatic contamination, we analyzed methylation at a maternally imprinted gene, DLK1, previously shown (Jenkins *et al.*, 2018) to be differentially methylated between sperm and somatic tissues as well as the paternally imprinted locus, H19 (Supplementary Fig. S1). At the DLK1 and H19 loci, average methylation for all participants was below 5%, and above 90%, respectively, suggesting negligible somatic cell contamination.

The first step in constructing our sperm epigenetic clock was to identify individual CpGs and differentially methylated regions (DMRs) that were significantly associated with male age. To facilitate the biological interpretation of our results, we reported associational estimates using β-values, which provide CpG methylation values between 0% and 100%. However, to identify age-associated sperm DNA methylation, we converted β-values to *M*-values, which is the logit transformation of the β-values: log (β/(1−β)), owing to its better adherence to homoscedasticity in linear models (Du *et al.*, 2010).

For individual CpGs, associations between male age and sperm methylation were determined using CpGassoc (Barfield *et al.*, 2012). Next, regional methylation analyses were employed via both the unsupervised adjacent clustering (A-clust) (Sofer *et al.*, 2013; Oluwayiose *et al.* 2022) and the supervised DMRcate algorithms (Peters *et al.*, 2015) to identify co-regulated CpGs (≥2 correlated CpG sites) within 1000 bp. For A-clust, a 0.25 Spearman correlation threshold of contiguous CpG sites within a given cluster was specified; a total of 21 872 co-regulated regions were identified and formed the unit of our regional methylation analyses. A general estimated equations model was employed to identify age-associated CpG regions by using a linear link and an exchangeable correlation to account for the correlated structure of the CpGs sites within a region (Liang and Zeger, 1986). For DMRcate, however, the regional clustering of individually significant probes (false discovery rate <0.05) was achieved using Gaussian kernel smoother with a scaling factor input of 2. These analyses were not adjusted for covariates; however, *P*-values were adjusted for multiple testing either by Benjamini–Hochberg ($q < 0.05$) or Bonferroni correction.

Sperm epigenetic aging clock

For the sperm epigenetic age clock via DMRs (SEA_{DMR}), sperm DMRs that were significantly associated with age and common between the two clustering approaches were used to predict chronological age via Super Learner (van der Laan et al., 2007), an ensemble machine learning method. Super Learner uses sample-splitting to combine predictions from multiple algorithms and thereby improve predictive performance. In this application, we considered penalized regressions (LASSO, ridge, elastic net) and multivariate adaptive regression splines (Friedman, 1991; Fan and Li, 2001; Hastie et al., 2009) with and without screening based on sure independence screening, LASSO, and elastic net penalties. These algorithms were selected to accommodate high dimensional data, while exploring both linear and non-linear relations between chronological age and sperm DMRs. The best weighted combination of algorithm-specific predictions was determined by minimizing the cross-validated (10-folds), mean-squared error. As performance metrics, mean absolute error (MAE) and correlations between true and predicted age were examined. To evaluate the out-of-sample performance, another layer of 10-fold cross-validation was added so that data used for training were distinct from data used for validation (Supplementary Fig. S2).

Utilizing an analogous approach, a sperm epigenetic age clock via individual CpGs (SEA_{CpG}) was constructed by applying age-associated individual CpGs as predictors into Super Learner. For both clocks, SEA was then calculated as the residuals from a linear regression of the Super Learner's predicted age on chronological age (Hannum et al., 2013). Positive values of SEA were considered an older epigenetic aging phenotype, while negative values represented a younger epigenetic aging phenotype. In sensitivity analyses, both clocks were constructed using β-value scales as predictors.

Ontology analyses

Prior to ontology analyses, each methylation region was assigned the closest gene within 1500 bp of the transcriptional start site (TSS) using GRCh37 assembly data from ENSEMBL via the annotatePeakInBatch function from the ChIPpeakAnno R package (version 3.6.5). We then used metascape (http://metascape.org) (Tripathi et al., 2015) to determine the functional enrichment of age-associated sperm DMRs in known biological processes and pathways.

Associational analyses

SEA was categorized into three groups for each clock: 'younger' for SEA < -1 year; 'equivalent' for SEA within 1 year, and 'older' for SEA >1 year. The Cochran–Armitage test was used to examine trends in non-mutually exclusive pregnancy categories at 3, 6 and 12 months across SEA groups. Additionally, unadjusted Kaplan–Meier survival analysis curves of TTP curves were generated to show differences in the SEA grouping on the probability of pregnancy across menstrual cycles. Then a discrete-time Cox proportional hazards model was used to assess the association between SEA as a continuous variable and TTP. Associations between SEA and continuous birth outcomes (GA, offspring birthweight, length, head circumference and ponderal index) and live birth were performed using multivariable linear and logistic regressions, respectively. All associational models were adjusted for male BMI (kg/m^2) and active smoking status and female chronological age, BMI and active smoking status. BMI was measured by research staff at the baseline interview and active smoking status was determined by a serum cotinine concentration of 10 ng/ml or greater (Bernert et al., 1997). SEA values were also summarized by quartiles of GA and ANOVA was used to test for trend and for equivalence of means. Finally, means and distributions of SEA by smoking status were compared using a Student's *t*-test and Kolmogorov–Smirnov test statistics, respectively. Analyses were run in R version 4.1.0 (R-Project, 2021).

Results

Our first step in constructing a biological clock for SEA was to identify DNA methylation at individual CpGs and regionally (i.e. clustered CpGs) that was significantly associated with chronological age among the 379 men recruited from the general population as part of the LIFE Study. Men were, on an average ±SD, 31.8 ± 4.8 years of age with a range of 19–50 years (Fig. 1A). Participants' demographics along with semen and reproductive outcomes are presented in Table I. Male participants were predominantly white (81.4%) men, non-smokers (78.7%) with a mean BMI of 29.9 ± 5.7. Among the 308 couples followed for 12 cycles, 150 (48.7%) became pregnant within the first three cycles of trying and 273 (88.6%) within 12 cycles; 35 (11.4%) couples did not have an observed pregnancy (data not shown). For the 192 couples with at least 21 weeks of GA and whose pregnancy resulted in live birth, 27 (14%) couples had preterm births (<35 weeks from date of ovulation). There were 46 participants without live birth.

Male age was associated with 85 434 CpGs ($q < 0.05$; Supplementary Table SI) and 22 397 CpGs after Bonferroni correction. Since regional methylation likely has more pronounced effects on gene expression than methylation at individual CpGs, we next identified age-associated sperm DMRs. To increase the rigor of our analyses, we employed two separate CpG clustering algorithms for regional methylation analyses, Aclust and DMRcate, and only included overlapping sperm DMRs in downstream analyses. In unadjusted linear models, 12 247 and 12 935 sperm DMRs were identified in Aclust and DMRcate analyses, respectively ($q < 0.05$), of which 2364 (Supplementary Table SII) sperm DMRs overlapped between the two analyses (Fig. 1B). Of these 2364 sperm DMRs, 1047 (44.3%) and 1317 (55.7%) were positively and negatively associated with age, respectively (Fig. 1C). Ontology analyses revealed that the overlapping sperm DMRs were enriched in numerous biological processes including multiple terms related to signaling, morphogenesis, brain development and learning or memory (Fig. 1D).

Next, we developed two independent biological clocks by applying Super Learner (van der Laan et al., 2007) to predict chronological age from the previously identified 22 397 Bonferroni age-associated individual CpGs and the 2364 age-associated sperm DMRs ($q < 0.05$) and then calculating SEA, as described in methods above.

Individual CpG-based SEA

Super Learner combines prediction algorithms with further screening algorithms. The resulting SEA_{CpG} utilized 120 sites (Supplementary Table SIII) with almost equivalent hypermethylated (49.2%) and hypomethylated (50.8%) sites. These CpGs were specifically located around

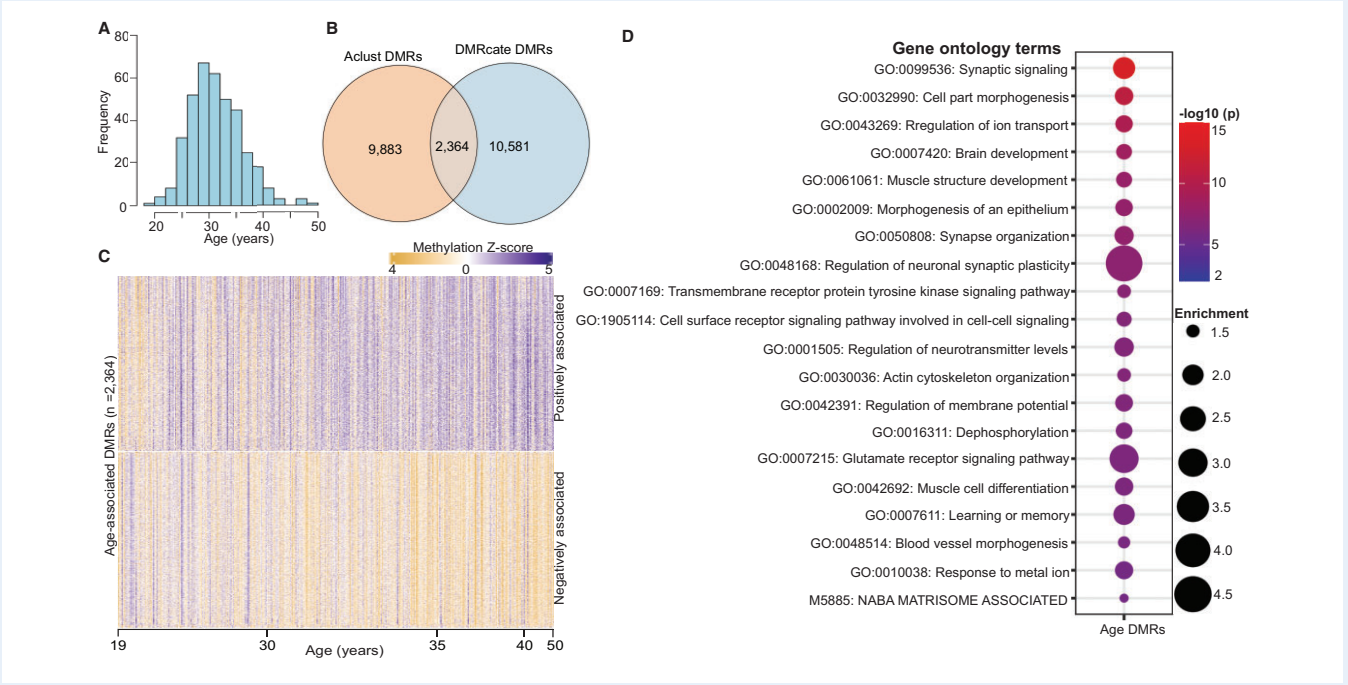

Figure 1. Association between chronological age and sperm DNA methylation. (**A**) Distribution of chronological age. (**B**) Venn diagram of age-associated differentially methylated regions (DMRs) from Aclust and DMRcate algorithms and their overlap. (**C**) Heatmap of the methylation profiles of the overlapping DMRs between Aclust and DMRcate. Purple and orange colors represent age-associated hypermethylated and hypomethylated DMRs. (**D**) Dotplot of gene ontology analysis of the shared DMRs between Aclust and DMRcate.

83 unique genes, which were enriched in top ontology terms including flavonoid glucuronidation, transmission across chemical synapses and histone ubiquitination (Fig. 2A). The annotation of SEA_{CpG} showed that the CpGs were enriched in CpG island shores and intergenic regions and depleted in CpG islands, exons and regions of known nucleosome retention in mature sperm (Supplementary Fig. S3). Our SEA_{CpG} model resulted in high predictive performance of chronological age within sample ($r = 0.91$, MAE = 1.6) (Fig. 2B) and out of sample ($r = 0.81$, MAE = 2.2) (Fig. 2C) and generated SEA_{CpG} that ranged from −4.4 to 5.8 years (Fig. 2D).

DMR-based SEA

To capture genomic regions with potentially more age-related biological relevance, we also applied Super Learner to genomic regions showing strong correlation with aging. Super Learner's prediction utilized 117 age-associated sperm DMRs (Supplementary Table SIV), more than half (62.39%) of which showed age-related hypomethylated sites that comprised 318 CpGs. Of these, 86 DMRs were located within 1500 bp of TSS of unique genes that were associated with top ontology terms including microtubule nucleation, maintenance of protein location in cell and regulation of small GTPase mediated signal transduction (Fig. 2E). The annotation of SEA_{DMR} showed that the DMRs were enriched in CpG island shores and shelves, and depleted in CpG islands, exons and open sea regions (Supplementary Fig. S3). The predictive performance of the SEA_{DMR} was high, yielding a strong association between chronological age with predicted age (correlation = 0.89) and a MAE of 1.7 years (Fig. 2F). Out-of-sample performance was similar: the cross-validated correlation was 0.79, and cross-validated MAE was 2.3 years (Fig. 2G). SEA_{DMR} values ranged from −4.4 to 6.3 years (Fig. 2H).

Sperm epigenetic age and reproductive outcomes

We were next interested in determining the utility of SEA as a novel biomarker of fecundity (TTP) and fertility (birth). We first examined relations by categorizing SEA in three groups: younger SEA (< −1 year); equivalent SEA (within ±1 year); older SEA (>1 year). In unadjusted models, we observed inverse associations with increasing SEA and a lower probability of achieving pregnancy within 3, 6 and 12 months of trying (Fig. 3A and C; categories not mutually exclusive). Based on SEA_{CpG}, the pregnancy probabilities within 12 months (the cutoff for diagnosis of clinical infertility), were 0.94 (95% CI: 0.86, 0.98) for younger SEA, 0.89 (95% CI: 0.83, 0.93) for equivalent SEA and 0.77 (95% CI: 0.65, 0.85) for older SEA, representing a 17% lower cumulative probability for couples with male partners in the older compared to the younger SEA_{CpG} categories (Fig. 2A). Similarly, for SEA_{DMR} the pregnancy probabilities at 12 months were 0.93 (95% CI: 0.84, 0.97); 0.93 (95% CI: 0.88, 0.96) and 0.70 (95% CI: 0.59, 0.80) for the younger, equivalent and older SEA_{DMR} groups, respectively (Fig. 2C), indicating a 23% lower probability of pregnancy in the older versus younger SEA_{DMR} category. Tests for trend strongly suggest a dose–response, with decreasing pregnancy probability for each time point with increasing SEA categories ($P < 0.05$; for all analyses).

Table I Demographic and reproductive characteristics of participants in the LIFE Study (n = 379).

Demographics by sex	Male	Female
Age (years): mean (sd)	31.7 (4.8)	29.9 (4.1)
BMI (kg/m^2)[1,2]: mean (sd)	29.9 (5.7)	27.4 (7.5)
Current smoking[3,4]: count (%)		
Yes	80 (21.3)	40 (10.6)
No	295 (78.7)	331 (87.4)
Race[5]: count (%)		
Non-Hispanic White	307 (81.4)	–
Non-Hispanic Black	15 (4.0)	–
Hispanic	32 (8.5)	–
Other	23 (6.1)	–
Baseline semen parameters	**Median (range)**	**n (%) < WHO cutoff**
Sperm concentration (x10^6/mL)	60.9 (2.0–332.8)	29 (7.7)
Normal morphology (%)[6]	30 (2.0–60.5)	4 (1.1)
Sperm count (10^6/ejaculate)	181.1 (3.5–1,141.2)	28 (7.4)
Semen volume (mL)	3.2 (0.4–9.4)	33 (8.7)
Reproductive outcomes	**Count (%)**	
TTP (cycles) (n =372)[7]		
≤ 3	150 (48.7%)	
≤ 6	237 (76.9%)	
≤ 12	273 (88.6%)	
Loss to follow-up	64 (17.2%)	
Live birth (n=238)[8]		
Yes	192 (70.3)	
No	46 (16.8)	
Gestational age (days) (n =192)[9]		
Preterm[10]	27 (14.1)	
Term	165 (85.9)	

– Not reported; *Total motility was measured after semen samples were shipped overnight;
[1]Missing for males (n = 2);
[2]Missing for females (n = 1);
[3]Missing for males (n = 4);
[4]Missing for females;
[5]Missing n = 2;
[6]Missing n = 24;
[7]Categories of time-to-pregnancy (TTP) are not mutually exclusive;
[8]Missing data for n = 33;
[9]Considered only participants (n =192) with at least 133 days (21 weeks) of gestational age (GA); and
[10]Preterm delivery outcome is defined as GA < 245 days (35 weeks) from the date of ovulation (conventionally defined as 37 weeks from last menstrual period date in the absence of ovulation data. LIFE: the Longitudinal Investigation of Fertility and the Environment Study, time-to-pregnancy (TTP); World Health Organization (WHO).

Unadjusted Kaplan–Meier curves further illustrated the dose-dependent relation between cycle-specific probabilities of pregnancy and categories of SEA for both clocks (Fig. 3B and D). Similarly, in adjusted Cox models, SEA was negatively associated with TTP (fecundability odds ratios (FORs): 0.83; 95% CI: 0.76–0.90; $P = 1.2\times10^{-5}$) for SEA_{CpG} and 0.85 (95% CI: 0.79–0.92; $P = 7.4\times10^{-5}$) for SEA_{DMR}. This indicates up to 17% lower fecundability, which corresponds to 17% longer TTP for every 1-year increase in SEA_{CpG}.

We also examined associations between SEA and birth outcomes. While no significant associations were found for most of the birth outcomes, including probability of a live birth, the SEA for both clocks was inversely associated with GA (Table II). Specifically, after controlling for male BMI and smoking status as well as female chronological age, BMI and smoking status, GA was ~2 days shorter for every increase in SEA: −2.13 days (95% CI: −3.67, −0.59) in our SEA_{CpG} clock and −1.89 (95% CI: −3.38, −0.41) in our SEA_{DMR} clock.

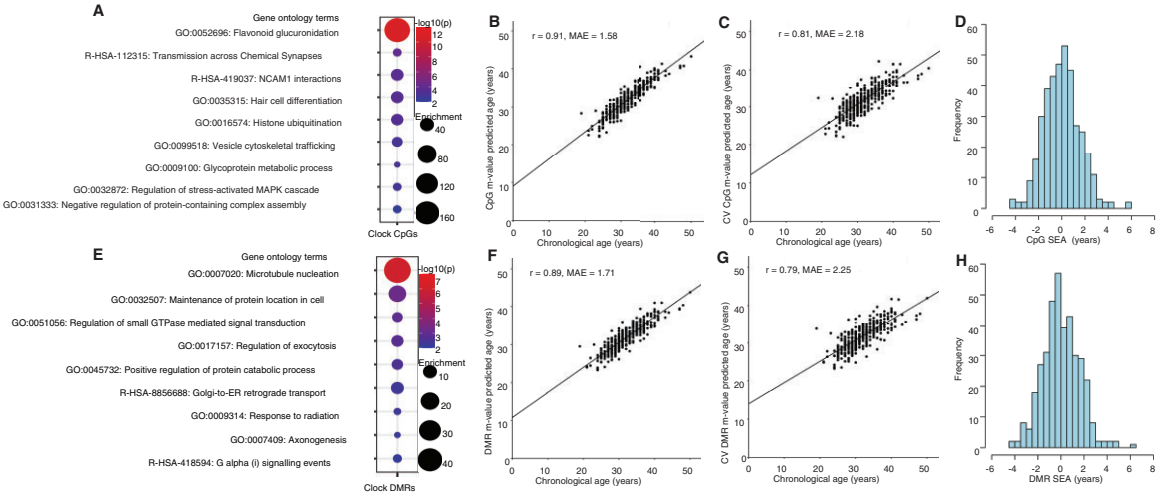

Figure 2. Super learner selected age-associated sperm methylation for epigenetic clock prediction from CpG and DMR-based clocks. (**A** and **E**) Dotplot of genes associated with the selected methylation loci. (**B** and **F**) Scatterplots of correlation between chronological and predicted age. (**C** and **G**) Scatterplots of correlation between chronological and out-of-sample (a.k.a., cross-validated; CV) predicted age. (**D** and **H**) Histogram of the distribution of sperm epigenetic aging (SEA). DMR, differentially methylated regions; MAE, mean absolute error.

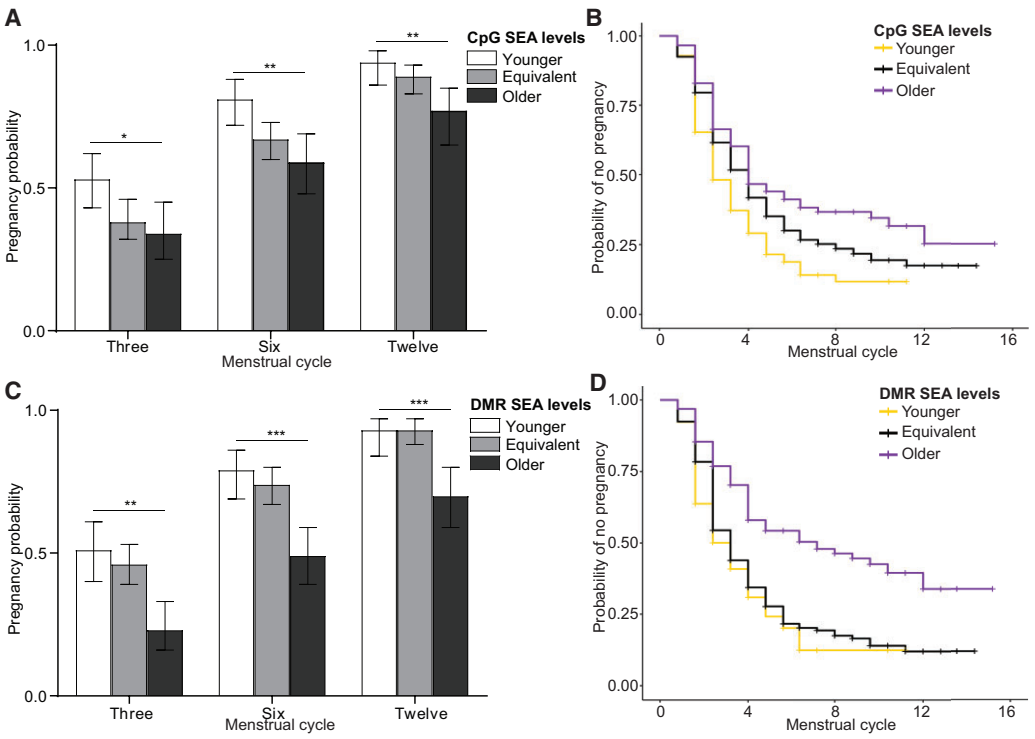

Figure 3. Association between sperm epigenetic aging and time-to-pregnancy from CpG and DMR-based clocks. (**A** and **C**) Barplot of cumulative probability of achieving pregnancy at 3, 6 and within 12 months. (**B** and **D**) Kaplan–Meier curves of the relationship between cycle-specific probabilities of pregnancy by levels of SEA. SEA, sperm epigenetic age; DMR, differentially methylated regions. *$P<0.05$; **$P<0.01$; and ***$P<0.0001$ as determined by Cochran–Armitage test for trend.

This inverse relationship was also supported by analysis of quartiles of GA across the two clocks (P-trend = 0.03 for both; Supplementary Fig. S4). Analyses of SEA with preterm (n = 27) versus normal term pregnancies (n = 165) were not significant, which may be attributable to the sample size of the preterm group.

Furthermore, we evaluated the influence of cigarette smoking on SEA. As expected, in both clocks, the average SEA±SD was higher among current smokers than non-smokers: SEA_{CpG} clock (0.42 ± 1.64 versus −0.12 ± 1.52, P = 0.008) and SEA_{DMR} clock (0.34 ± 1.78 versus −0.10 ± 1.53, P = 0.04; Fig. 4A and B).

While we were unable to obtain a second semen sample from LIFE participants to evaluate the internal reproducibility of our results, we evaluated the external generalizability of our approach by using Super Learner, trained on the LIFE cohort (n = 379), to predict chronological age among participants in SEEDS, a new independent clinical cohort of men (n = 173) enrolled for infertility treatment. We found a strong correlation between predicted and chronological age (CpG-based model: r = 0.83 and DMR-based model: r = 0.79; Supplementary Fig. S5A and B), suggesting the potential utility of our models across population groups.

Discussion

Chronological age is a significant determinant of reproductive capacity and success among couples (Dunson *et al.*, 2002; Mutsaerts *et al.*, 2012; Sharma *et al.*, 2015; Oluwayiose *et al.*, 2021). However, chronological age does not encapsulate the cumulative internal (e.g. genetics) and external (e.g. environmental conditions) factors that ensue over the life-course, and thus it serves as a proxy measure of the 'true' biological age of cells. While semen quality outcomes utilizing WHO guidelines (Cooper *et al.*, 2010) have been used for the assessment of male infertility for decades, they remain poor predictors of reproductive outcomes (Buck Louis *et al.*, 2011; Jungwirth *et al.*, 2012). Thus, the ability to capture the biological age of sperm may provide a novel platform to better assess the male contribution to reproductive success, especially among idiopathic infertile couples. Here, we report, for the first time to our knowledge, that our novel approach to estimate SEA is efficacious in strongly predicting couple fecundity, as measured by TTP, among couples from the general population. Our results indicate that higher SEA (both continuous and categorical) is associated with a longer TTP as well as shorter gestation among couples becoming pregnant. SEA was also higher among males who were current cigarette smoking. These results among pregnancy planners in the general population who are not seeking clinical fertility treatment are novel and hold promise to overcome the limitations of using conventional semen quality in predicting couples' reproductive outcomes.

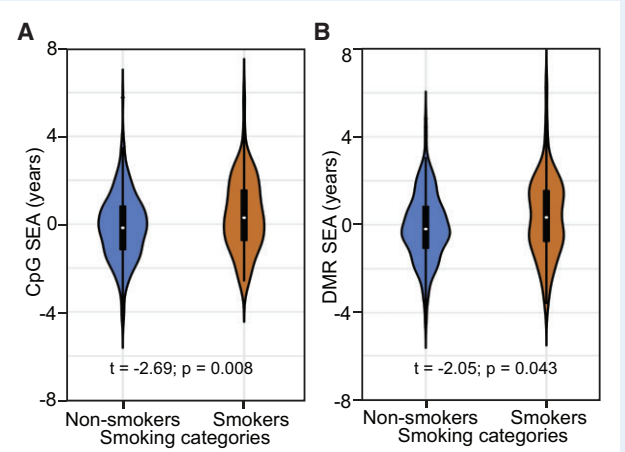

Figure 4. Comparison of SEA between smokers and non-smokers. Violin plots of (**A**) individual CpG and (**B**) DMR clocks. DMR, differentially methylated regions; SEA, sperm epigenetic aging.

Table II Associations between sperm epigenetic aging and reproductive and birth outcomes.

Outcomes	CpG SEA		DMR SEA	
	Estimate (CI)	P-value	Estimate (CI)	P-value
TTP (FOR)[1]	0.83 (0.76, 0.90)	1.2×10^{-5}	0.85 (0.79, 0.92)	7.4×10^{-5}
Gestational age (days)[2]	−2.13 (−3.67, −0.59)	0.005	−1.89 (−3.38, −0.41)	0.01
Birthweight (lb)[2]	−0.02 (−0.12, 0.09)	0.74	−0.05 (−0.16, 0.05)	0.30
Birth length (in)[2]	−0.001 (−0.11, 0.11)	0.98	−0.05 (−0.16, 0.05)	0.34
Head circumference (in)[2]	−0.04 (−0.13, 0.06)	0.43	−0.05 (−0.14, 0.04)	0.21
Ponderal index (100xBL/BW^3)[2]	−0.01 (−0.04, 0.02)	0.66	−0.00 (−0.03, 0.03)	0.99
Live birth (OR)[3]	0.88 (0.69, 1.12)	0.31	0.92 (0.73, 1.17)	0.50

All models were adjusted for male BMI and smoking status and female chronological age, BMI and smoking status.
[1]n = 372;
[2]n = 192;
[3]n = 238.
SEA, sperm epigenetic aging; DMR, differentially methylated regions; TTP, time-to-pregnancy; FOR, fecundability odds ratio; lb, pounds; in, inches; BL, birth length; BW, birthweight; OR, odds ratio.
TTP analyses were conducted via discrete-time Cox proportional hazards models, while birth outcomes and live birth were performed using multivariable linear and logistic regressions, respectively.

In regard to epigenetic aging of sperm, we must consider the developmental stage whereby age has its greatest influence on DNA methylation patterns overtime. Male germ cells require dynamic epigenetic reprogramming for the progression from diploid spermatogonia to haploid spermatozoa (Godmann *et al.*, 2009; Marcho *et al.*, 2020). Although it is recognized that the final DNA methylation patterns are established during meiotic divisions (Oakes *et al.*, 2007), the accumulation of aging-related methylation errors likely occur in the highly proliferative and self-renewing spermatogonia and are carried forward when cells are committed to differentiation during spermatogenesis. As such, SEA likely reflects that of spermatogonia, from which mature spermatozoa are generated during the 74-day process of spermatogenesis. In humans, it is estimated that spermatogonia divide every 16 days, which equates to 23 divisions a year (Goriely, 2016). Thus, the spermatogonia of the oldest participant in our study of 50 years would have undergone (taking into account approximate age of puberty) >800 divisions during his reproductive life-course. A consensus for the genomic context (intergenic versus gene regions) of these accumulated methylation errors remains to be resolved. Detailed nucleosome maps as well as 3D/4D architecture and dynamics of spermatogonia, not mature spermatozoa, in space and time hold promise to uncover why certain regions are more susceptible to age-related epigenetic dysregulation.

Emerging data over the last few years demonstrate the profound effect of aging on the sperm methylome and their potential for constructing epigenetic clocks (Jenkins *et al.*, 2018; Cao *et al.*, 2020; Laurentino *et al.*, 2020; Oluwayiose *et al.*, 2021); however, the clinical relevance of these clocks has remained largely unexplored. Among 329 samples from infertile patients, sperm donors and individuals from the general population, Jenkins *et al.* (2018) observed that sperm methylation at 51 genomic regions (via Illumina's 450K) reproducibly predicted an individual's chronological age regardless of fertility status ($r = 0.89$; MAE = 2.04). Similar to our results, smokers had higher SEA compared to never smokers. Comparing young (n = 6; 18–24 years) and old men (n = 6; 61–71 years) men, Laurentino et al. identified 236 age-related sperm DMRs via shotgun sequencing (Laurentino *et al.*, 2020). Six DMRs with the lowest *P*-value were selected to build an epigenetic clock in 42 additional samples and subsequently in an independent set of 33 samples; however, the clock yielded high errors (MAEs = 7.8 and 9.8 years, respectively), which is likely attributable to the smaller sample sizes and small number of DMRs used in their analyses. Most recently, a customized methyl-capture sequencing approach identified 798 age-associated sperm DMRs by categorizing men as either young (n = 48; 18–38 years) or old (n = 46; 46–71 years) (Cao *et al.*, 2020). Elastic net analyses utilizing the top 5000 age-associated CpGs generated a sperm clock with an average error of 2.7 years. The authors note that their age prediction improved by increasing the number to thousands of CpGs; however, this puts into question the balance between assay efficiency (e.g. measuring thousands compared to 120 CpGs in our SEA_{CpG} clock) and incremental improvement in predictive value of SEA. A distinct advantage of our approach was to build a clock specifically to understand the impact of biological age on reproductive outcomes in the general population among couples who were discontinuing contraception for purposes of trying to become pregnant. Jenkins *et al.* (2018) combined samples from sperm donors and infertility patients; however, this approach may obscure differences in the biological aging patterns across these groups. Furthermore, the application of our clocks to predict biological age in an independent cohort of men seeking infertility treatment provided strong evidence of its relevancy for the general population.

Our epigenetic clocks are the first to employ Super Learner, which uses state-of-the-art machine learning methods to improve predictive performance. Previous clocks have relied on penalized linear regression (Jenkins *et al.*, 2018; Cao *et al.*, 2020; Laurentino *et al.*, 2020). This approach requires the strong assumption that chronological age is related to the DNA methylation in a simple linear fashion; in other words, there are no interaction terms or other non-linear effects. Additionally, the penalty terms in LASSO, elastic net or ridge regression aim to balance the total number of features with predictive performance; in so doing, some relevant features might be excluded to avoid overfitting. In contrast, our approach uses cross-validation to create an optimal weighted combination of multiple prediction algorithms, including both linear and non-linear approaches. Super Learner has improved performance in a variety of settings, including predicting mortality in intensive care units, violence in prisons and HIV risk in resource-limited settings (Pirracchio *et al.*, 2015; Baćak and Kennedy, 2019; Balzer *et al.*, 2020). Indeed, Super Learner is theoretically guaranteed to perform at least as well as the best algorithm in its ensemble (van der Laan and Dudoit, 2003). In this application, we considered both penalized regressions as well as multivariate adaptive regression splines, which is a highly non-linear and flexible approach. Our Super Learner SEA_{CpG} resulted in a low MAE of 1.6 and, as expected, outperformed prediction when relying solely on elastic net, which yielded a MAE of 2.2 (data not shown). Importantly, the weights assigned to each algorithm differed for the SEA_{CpG} and SEA_{DMR} clocks, underscoring the flexibility of the Super Learner algorithm.

Previous epigenetic clocks in somatic tissue (Hannum *et al.*, 2013; Horvath, 2013; Levine *et al.*, 2018) and sperm (Jenkins *et al.*, 2018; Cao *et al.*, 2020; Laurentino *et al.*, 2020) have relied on either individual CpGs or regional (DMR-based) approaches; however, the comparison of these two approaches is limited within the same study. In our approach, Super Learner utilized 22 397 Bonferroni age-associated CpGs and over 12 000 DMRs ($q < 0.05$) and selected 120 CpGs and 117 DMRs (comprising 318 CpGs) for our SEA_{CpG} and SEA_{DMR} clocks, respectively. In terms of out-of-sample performance, both SEA_{CpG} and SEA_{DMR} clocks performed well, yielding high accuracy of prediction and low error ($r = 0.81$; MAE = 2.2 years and $r = 0.79$; MAE = 2.3 years, respectively; all metrics cross-validated). Surprisingly, we found minimal overlap of methylation sites between our SEA_{CpG} and SEA_{DMR} clocks, such that only 10 CpG (<1%) of individual CpGs were present in the 318 CpGs of the 117 DMRs. This suggests that the two clocks harbor methylation sites in distinct genomic regions and may have independent utility for downstream reproductive outcome analyses. However, while both clocks were significantly associated with TTP, GA and male smoking, our SEA_{CpG} clock generated larger effect sizes in both linear and categorical analyses. Moreover, in regard to a practical application, our SEA_{CpG} clock requires ~ 3x less CpGs compared to our SEA_{DMR} clock, and thus is more efficient and would minimize the cost of constructing a custom methylation array without any discernable differences in performance. Although it is recognized that regional methylation status may have more profound effects on downstream gene expression, our data suggest that our SEA_{CpG} clock is the preferred approach to estimate SEA to predict reproductive outcomes over our SEA_{DMR} approach.

The strong and consistent relation between chronological age and DNA methylation at specific loci across individuals indicates that these age-associated changes in DNA methylation are likely not stochastic, but rather they occur at targeted regions that are more prone to epigenetic errors. The annotation of our CpG clock shows that the CpGs are enriched in CpG shores and intergenic regions, while depleted in CpG islands, exons and regions of known nucleosome retention in mature sperm. Taken together, our clock's CpGs are distal from genes and regions of nucleosome retention, which have been linked with genes important for embryo development (Arpanahi *et al.*, 2009; Hammoud *et al.*, 2009). Our previous research has shown that sperm DNA methylation mediated the effect of male chronological age on poor reproductive outcomes such as fertilization rates, embryo development and live birth (Oluwayiose *et al.*, 2021). Interestingly, consistent with our results here, we found that age-associated CpGs were depleted in retained nucleosome regions, suggesting that the effect of age via sperm DNA methylation on reproductive outcomes, such as TTP, may be independent of genic regions known to retain nucleosomes (Oluwayiose *et al.*, 2021). Other groups have reported that age-related hypomethylated DMRs were enriched in gene regions, while hypermethylated DMRs were enriched in distal regions (Cao *et al.*, 2020); however, we observed no such distinction upon stratification of our clock CpGs by change in methylation direction.

While our study utilized gradient centrifugation to remove somatic cell contamination, future studies could employ flow cytometry to further isolate haploid sperm from leukocyte and non-haploid sperm contamination. Owing to the strong association between SEA and couples' pregnancy probability, the slowing or reversal of SEA through lifestyle choices and/or pharmacological interventions warrants further investigation. Therefore, the characterization of the potential pathways driving the losses/gains of methylation with age offers innovative avenues of translational research to mitigate sperm aging. Moreover, as older fathers have an increased risk of offspring with a host of adverse neurological outcomes (Montgomery *et al.*, 2004; Saha *et al.*, 2009; Puleo *et al.*, 2012), it is of clinical importance to understand the potential relation of SEA on offspring health and development, such as our significant findings with GA, and if it is a more precise predictor of risk of adverse offspring health.

Conclusion

There is a critical need for new measures of male fecundity for assessing overall reproductive success among couples in the general population. These data show that our sperm epigenetic clocks may fulfill this need as a novel biomarker that predicts pregnancy success among couples not seeking fertility treatment. While chronological age of both partners remains a significant predictor of reproductive success, our clocks, generated with as few as 120 CpG sites, likely recapitulates both external and internal factors that drive the biological aging of sperm. Such a summary measure of sperm biological age is of clinical importance as it allows couples in the general population to realize their probability of achieving pregnancy during natural intercourse, thereby informing and expediting potential infertility treatment decisions. With the ability to customize high throughput DNA methylation arrays and capture sequencing approaches, the integration of our epigenetic clocks as part of standard clinical care can enhance our understanding of idiopathic infertility and the paternal contribution to reproductive success and offspring health.

Supplementary data

Supplementary data are available at *Human Reproduction* online.

Data availability

DNA methylation data will be available at the GEO repository upon publication at GSE185445.

Authors' roles

J.R.P. oversaw study design and sperm epigenetic data, interpreted results and drafted the manuscript. H.S. performed analyses pertaining to the clocks and interpreted results. B.W.W. oversaw study design and statistical analyses and interpreted results. A.S. oversaw sperm epigenetic data and interpreted results. G.M.B.L. was the principal investigator for the LIFE Study and helped design and interpret the results presented in this paper. O.A.O. conducted all sperm epigenetic analyses and interpreted results. L.B.B. oversaw analyses pertaining to the clocks, interpreted results and edited the manuscript.

Funding

This work was funded in part by grants the National Institute of Environmental Health Sciences, National Institutes of Health (R01 ES028298; PI: J.R.P. and P30 ES020957); Robert J. Sokol, MD Endowed Chair of Molecular Obstetrics and Gynecology (J.R.P.); and the Intramural Research Program of the Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, Maryland (Contracts N01-HD-3-3355, N01-HD-3-3356 and N01-HD-3-3358). S.L.M. was supported by the Intramural Research Program of the Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health.

Conflict of interest

The authors declare no competing interests.

References

Arpanahi A, Brinkworth M, Iles D, Krawetz SA, Paradowska A, Platts AE, Saida M, Steger K, Tedder P, Miller D. Endonuclease-sensitive regions of human spermatozoal chromatin are highly enriched in promoter and CTCF binding sequences. *Genome Res* 2009;**19**: 1338–1349.

Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, Irizarry RA. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* 2014;**30**:1363–1369.

Baćak V, Kennedy EH. Principled machine learning using the super learner: an application to predicting prison violence. *Sociol Methods Res* 2019;**48**:698–721.

Balzer LB, Havlir DV, Kamya MR, Chamie G, Charlebois ED, Clark TD, Koss CA, Kwarisiima D, Ayieko J, Sang N *et al*. Machine learning to identify persons at high-risk of human immunodeficiency virus acquisition in rural Kenya and Uganda. *Clin Infect Dis* 2020;**71**: 2326–2333.

Barfield RT, Kilaru V, Smith AK, Conneely KN. CpGassoc: an R function for analysis of DNA methylation microarray data. *Bioinformatics* 2012;**28**:1280–1281.

Bernert JT Jr, Turner WE, Pirkle JL, Sosnoff CS, Akins JR, Waldrep MK, Ann Q, Covey TR, Whitfield WE, Gunter EW *et al*. Development and validation of sensitive method for determination of serum cotinine in smokers and nonsmokers by liquid chromatography/atmospheric pressure ionization tandem mass spectrometry. *Clin Chem* 1997;**43**:2281–2291.

Bloom MS, Whitcomb BW, Chen Z, Ye A, Kannan K, Buck Louis GM. Associations between urinary phthalate concentrations and semen quality parameters in a general population. *Hum Reprod* 2015;**30**:2645–2657.

Boivin J, Bunting L, Collins JA, Nygren KG. International estimates of infertility prevalence and treatment-seeking: potential need and demand for infertility medical care. *Hum Reprod* 2007;**22**:1506–1512.

Breitling LP, Saum K-U, Perna L, Schöttker B, Holleczek B, Brenner H. Frailty is associated with the epigenetic clock but not with telomere length in a German cohort. *Clin Epigenet* 2016;**8**.

Buck Louis GM, Schisterman EF, Sweeney AM, Wilcosky TC, Gore-Langton RE, Lynch CD, Boyd Barr D, Schrader SM, Kim S, Chen Z *et al*. Designing prospective cohort studies for assessing reproductive and developmental toxicity during sensitive windows of human reproduction and development–the LIFE Study. *Paediatr Perinat Epidemiol* 2011;**25**:413–424.

Buck Louis GM, Sundaram R, Schisterman EF, Sweeney A, Lynch CD, Kim S, Maisog JM, Gore-Langton R, Eisenberg ML, Chen Z. Semen quality and time to pregnancy: the Longitudinal Investigation of Fertility and the Environment Study. *Fertil Steril* 2014;**101**:453–462.

Cao M, Shao X, Chan P, Cheung W, Kwan T, Pastinen T, Robaire B. High-resolution analyses of human sperm dynamic methylome reveal thousands of novel age-related epigenetic alterations. *Clin Epigenetics* 2020;**12**:192.

Cooper TG, Noonan E, von Eckardstein S, Auger J, Baker HW, Behre HM, Haugen TB, Kruger T, Wang C, Mbizvo MT *et al*. World Health Organization reference values for human semen characteristics. *Hum Reprod Update* 2010;**16**:231–245.

Degerman S, Josefsson M, Nordin Adolfsson A, Wennstedt S, Landfors M, Haider Z, Pudas S, Hultdin M, Nyberg L, Adolfsson R. Maintained memory in aging is associated with young epigenetic age. *Neurobiol Aging* 2017;**55**:167–171.

Du P, Zhang X, Huang CC, Jafari N, Kibbe WA, Hou L, Lin SM. Comparison of beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics* 2010; **11**:587.

Dunson DB, Colombo B, Baird DD. Changes with age in the level and duration of fertility in the menstrual cycle. *Hum Reprod* 2002; **17**:1399–1403.

Fan J, Li R. Variable selection via nonconcave penalized likelihood and its oracle properties. *J Am Stat Assoc* 2001;**96**:1348–1360.

Fraga MF, Ballestar E, Paz MF, Ropero S, Setien F, Ballestar ML, Heine-Suner D, Cigudosa JC, Urioste M, Benitez J *et al*. From the cover: epigenetic differences arise during the lifetime of monozygotic twins. *Proc Natl Acad Sci U S A* 2005;**102**:10604–10609.

Fransquet PD, Wrigglesworth J, Woods RL, Ernst ME, Ryan J. The epigenetic clock as a predictor of disease and mortality risk: a systematic review and meta-analysis. *Clin Epigenet* 2019;**11**.

Friedman JH. Multivariate adaptive regression splines. *Ann Stat* 1991; **19**:1–141.

Godmann M, Lambrot R, Kimmins S. The dynamic epigenetic program in male germ cells: Its role in spermatogenesis, testis cancer, and its response to the environment. *Microsc Res Tech* 2009;**72**:603–619.

Goriely A. Decoding germline de novo point mutations. *Nat Genet* 2016;**48**:823–824.

Hammoud SS, Nix DA, Zhang H, Purwar J, Carrell DT, Cairns BR. Distinctive chromatin in human sperm packages genes for embryo development. *Nature* 2009;**460**:473–478.

Hannum G, Guinney J, Zhao L, Zhang L, Hughes G, Sadda S, Klotzle B, Bibikova M, Fan JB, Gao Y *et al*. Genome-wide methylation profiles reveal quantitative views of human aging rates. *Mol Cell* 2013; **49**:359–367.

Hassan MAM, Killick SR. Effect of male age on fertility: evidence for the decline in male fertility with increasing age. *Fertil Steril* 2003;**79**: 1520–1527.

Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer Series in Statistics*, 2nd edn. New York: Springer-Verlag, 2009.

Hemminki K, Kyyrönen P. Parental age and risk of sporadic and familial cancer in offspring: implications for germ cell mutagenesis. *Epidemiology* 1999;**10**:747–751.

Horta F, Vollenhoven B, Healey M, Busija L, Catt S, Temple-Smith P. Male ageing is negatively associated with the chance of live birth in IVF/ICSI cycles for idiopathic infertility. *Hum Reprod* 2019;**34**: 2523–2532.

Horvath S. DNA methylation age of human tissues and cell types. *Genome Biol* 2013;**14**:R115.

Jacobsson B, Ladfors L, Milsom I. Advanced maternal age and adverse perinatal outcome. *Obstet Gynecol* 2004;**104**:727–733.

Jenkins TG, Aston KI, Cairns B, Smith A, Carrell DT. Paternal germ line aging: DNA methylation age prediction from human sperm. *BMC Genomics* 2018;**19**:763.

Jenkins TG, Aston KI, Carrell DT. Sperm epigenetics and aging. *Transl Androl Urol* 2018;**7**:S328–S335.

Jungwirth A, Giwercman A, Tournaye H, Diemer T, Kopa Z, Dohle G, Krausz C; European Association of Urology Working Group on Male Infertility. European Association of Urology Guidelines on Male Fertility: the 2012 update. *Eur Urol* 2012;**62**:324–332.

Kemkes-Grottenthaler A. Parental effects on offspring longevity – evidence from 17th to 19th century reproductive histories. *Ann Hum Biol* 2004;**31**:139–158.

Landmann E, Reiss I, Misselwitz B, Gortner L. Ponderal index for discrimination between symmetric and asymmetric growth restriction: percentiles for neonates from 30 weeks to 43 weeks of gestation. *J Matern Fetal Neonatal Med* 2006;**19**:157–160.

Laurentino S, Cremers J-F, Horsthemke B, Tüttelmann F, Czeloth K, Zitzmann M, Pohl E, Rahmann S, Schröder C, Berres S *et al*. A germ cell-specific ageing pattern in otherwise healthy men. *Aging Cell* 2020;**19**:e13242.

Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The SVA package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* 2012;**28**: 882–883.

Levine ME, Lu AT, Chen BH, Hernandez DG, Singleton AB, Ferrucci L, Bandinelli S, Salfati E, Manson JE, Quach A *et al*. Menopause accelerates biological aging. *Proc Natl Acad Sci U S A* 2016;**113**: 9327–9332.

Levine ME, Lu AT, Quach A, Chen BH, Assimes TL, Bandinelli S, Hou L, Baccarelli AA, Stewart JD, Li Y *et al*. An epigenetic biomarker of aging for lifespan and healthspan. *Aging (Albany NY)* 2018;**10**:573–591.

Liang K-Y, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika* 1986;**73**:13–22.

Marcho C, Oluwayiose OA, Pilsner JR. The preconception environment and sperm epigenetics. *Andrology* 2020;**8**:924–942.

Montgomery SM, Olsson T, Ekbom A. Parental age, family size, and risk of multiple sclerosis. *Epidemiology* 2004;**15**:717–723.

Mutsaerts MA, Groen H, Huiting HG, Kuchenbecker WK, Sauer PJ, Land JA, Stolk RP, Hoek A. The influence of maternal and paternal factors on time to pregnancy–a Dutch population-based birth-cohort study: the GECKO Drenthe study. *Hum Reprod* 2012;**27**: 583–593.

Niu L, Xu Z, Taylor JA. RCP: A novel probe design bias correction method for Illumina Methylation BeadChip. *Bioinformatics* 2016;**32**: 2659–2663.

Oakes CC, La Salle S, Smiraglia DJ, Robaire B, Trasler JM. Developmental acquisition of genome-wide DNA methylation occurs prior to meiosis in male germ cells. *Dev Biol* 2007;**307**: 368–379.

Oluwayiose OA, Wu H, Saddiki H, Whitcomb BW, Balzer LB, Brandon N, Suvorov A, Tayyab R, Sites CK, Hill L *et al*. Sperm DNA methylation mediates the association of male age on reproductive outcomes among couples undergoing infertility treatment. *Sci Rep* 2021;**11**:3216.

Oluwayiose OA, Wu H, Gao F, Baccarelli AA, Sofer T, Pilsner JR. Aclust2.0: a revamped unsupervised R tool for Infinium methylation beadchips data analyses. 2022. https://github.com/OluwayioseOA/Aclust2.0.

Peng C, Cardenas A, Rifas-Shiman SL, Hivert M-F, Gold DR, Platts-Mills TA, Lin X, Oken E, Avila L, Celedón JC *et al*. Epigenetic age acceleration is associated with allergy and asthma in children in Project Viva. *J Allergy Clin Immunol* 2019;**143**: 2263–2270.e14.

Perna L, Zhang Y, Mons U, Holleczek B, Saum K-U, Brenner H. Epigenetic age acceleration predicts cancer, cardiovascular, and all-cause mortality in a German case cohort. *Clin Epigenetics* 2016; **8**:64.

Peters TJ, Buckley MJ, Statham AL, Pidsley R, Samaras K, V Lord R, Clark SJ, Molloy PL. De novo identification of differentially methylated regions in the human genome. *Epigenetics Chromatin* 2015;**8**:6.

Pirracchio R, Petersen ML, Carone M, Rigon MR, Chevret S, van der Laan MJ. Mortality prediction in intensive care units with the Super ICU Learner Algorithm (SICULA): a population-based study. *Lancet Respir Med* 2015;**3**:42–52.

Puleo CM, Schmeidler J, Reichenberg A, Kolevzon A, Soorya LV, Buxbaum JD, Silverman JM. Advancing paternal age and autism. *Autism* 2012;**16**:367–380.

R-Project. *A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing, 2021.

Saha S, Barnett AG, Foldi C, Burne TH, Eyles DW, Buka SL, McGrath JJ. Advanced paternal age is associated with impaired neurocognitive outcomes during infancy and childhood. *PLoS Med* 2009;**6**:e40.

Sharma R, Agarwal A, Rohra VK, Assidi M, Abu-Elmagd M, Turki RF. Effects of increased paternal age on sperm quality, reproductive outcome and associated epigenetic risks to offspring. *Reprod Biol Endocrinol* 2015;**13**:35.

Smith JF, Eisenberg ML, Millstein SG, Nachtigall RD, Sadetsky N, Cedars MI, Katz PP; Infertility Outcomes Program Project Group. Fertility treatments and outcomes among couples seeking fertility care: data from a prospective fertility cohort in the United States. *Fertil Steril* 2011;**95**:79–84.

Sofer T, Schifano ED, Hoppin JA, Hou L, Baccarelli AA. Gene expression A-clustering: a novel method for the detection of co-regulated methylation regions, and regions associated with exposure. *Bioinformatics* 2013;**29**:2884–2891.

Thoma ME, McLain AC, Louis JF, King RB, Trumble AC, Sundaram R, Buck Louis GM. Prevalence of infertility in the United States as estimated by the current duration approach and a traditional constructed approach. *Fertil Steril* 2013;**99**:1324–1331.e1.

Triche TJ, Weisenberger DJ, Van Den Berg D, Laird PW, Siegmund KD. Low-level processing of Illumina Infinium DNA Methylation BeadArrays. *Nucleic Acids Res* 2013;**41**:1–11.

Tripathi S, Pohl MO, Zhou Y, Rodriguez-Frandsen A, Wang G, Stein DA, Moulton HM, Dejesus P, Che J, Mulder LCF *et al*. Meta-and orthogonal integration of influenza "OMICs" data defines a role for UBR4 in virus budding. *Cell Host Microbe* 2015;**18**:723–735.

van der Laan MJ, Polley EC, Hubbard AE. Super learner. *Stat Appl Genet Mol Biol* 2007;**6**:Article25.

van der Laan MJ, Dudoit S. Unified Cross-Validation Methodology for Selection among Estimators and a General Cross-Validated Adaptive Epsilon-Net Estimator: Finite Sample Oracle Inequalities and Examples. *UC Berkeley Division of Biostatistics Working Paper Series*, 2003; Working Paper 130. https://biostats.bepress.com/ucbbiostat/paper130.

Waldenström U. Postponing parenthood to advanced age. *Ups J Med Sci* 2016;**121**:235–239.

Whitman-Elia GF, Baxley EG. A primary care approach to the infertile couple. *J Am Board Fam Pract* 2001;**14**:33–45.

WHO: recommended definitions, terminology and format for statistical tables related to the perinatal period and use of a new certificate for cause of perinatal deaths. Modifications recommended by FIGO as amended October 14, 1976. *Acta Obstet Gynecol Scand* 1977;**56**:247–253.

Wilcox AJ, Weinberg CR, Baird DD. Timing of sexual intercourse in relation to ovulation. Effects on the probability of conception, survival of the pregnancy, and sex of the baby. *N Engl J Med* 1995;**333**:1517–1521.

Wu H, de Gannes MK, Luchetti G, Pilsner JR. Rapid method for the isolation of mammalian sperm DNA. *BioTechniques* 2015;**58**: 293–300.

Wu H, Estill MS, Shershebnev A, Suvorov A, Krawetz SA, Whitcomb BW, Dinnie H, Rahil T, Sites CK, Pilsner JR. Preconception urinary phthalate concentrations and sperm DNA methylation profiles among men undergoing IVF treatment: a cross-sectional study. *Hum Reprod* 2017;**32**:2159–2169.