Tracking single cell evolution via clock-like chromatin accessibility

Xinghuan Wang[1,2,#,*], Wan Jin[3,#], Gang Wang[1,4,#], Lingao Ju[4,#], Fangjin Chen[5], Kaiyu Qian[1,4], Yu Xiao[2,4,*], Yi Zhang[3,*,†]

Affiliations:

[1]Department of Urology, Zhongnan Hospital of Wuhan University, Wuhan, China.

[2]Wuhan Research Center for Infectious Diseases and Cancer, Chinese Academy of Medical Sciences, Wuhan, China.

[3]Euler Technology, ZGC Life Sciences Park, Beijing, China.

[4]Department of Biological Repositories, Zhongnan Hospital of Wuhan University, Wuhan, China.

[5]High Performance Computing Center, the Peking-Tsinghua College of Life Sciences, Peking University, Beijing, China.

[#]These authors contributed equally to this work.

[†]Lead contact.

[*]Correspondence to Dr. Xinghuan Wang (wangxinghuan@whu.edu.cn), Dr. Yu Xiao (yu.xiao@whu.edu.cn) or Dr. Yi Zhang (zy@eulertechnology.com).

Running title: Clock-like chromatin accessibility tracks single cell mitotic age.

Abstract

Single cell chromatin accessibility sequencing (scATAC) reconstructs developmental trajectory by phenotypic similarity. However, inferring the exact developmental trajectory is challenging. Here we show a simple, accurate and phenotypic-neutral measure of cell developmental hierarchy – the fraction of accessible clock-like differential methylation loci (ClockDML). As cell undergone mitosis, heterogeneity of chromatin accessibility on ClockDML reduced, providing a measure of mitotic age. We developed a method, *EpiTrace*, that counts the fraction of opened ClockDML from scATAC data to determine cell age and perform lineage tracing. *EpiTrace* derived cell age shows concordance to known developmental hierarchies, correlates well with DNA methylation-based clocks, and is complementary with mutation-based lineage tracing, RNA velocity, and stemness predictions. Applying *EpiTrace* to human scATAC data revealed a multitude of novel biological insights with clinically relevant implications, ranging from hematopoiesis, organ development, tumor biology and immunity to cortical gyrification. Our work discovered a universal epigenomic hallmark during cellular development, which facilitates the study of cellular hierarchies and organismal aging.

Introduction

Single cell chromatin accessibility sequencing (scATAC) is a powerful technique for interrogating epigenomic landscape on single cell resolution[1]. However, inferring the exact developmental trajectory from scATAC data is challenging. Whilst excellent tools such as RNA velocity, stemness prediction and metabolic labelling[2-5] exist for determining cell evolution trajectory on the manifold of phenotype – usually described as Waddington's landscape – for single cell RNA-seq datasets, no comparable methods exist for scATAC. State-of-the-art, similarity-based lineage deduction methods[6] would be limited when phenotypes are fluidic such as in de-differentiation or oncogenesis. On the other hand, mutation-based lineage tracing methods, for example using mitochondrial SNPs[7-9], which tracks the phylogeny of cells over divisions, is highly accurate yet their temporal resolution is restrained by the low natural mutation rate.

The concept of mitotic age refers to the accumulative counts of mitosis a cell undergone after fertilization – or the ground state of cell division. Whilst the first proposed mitotic age biomarker – telomere length – was genetic[10,11], the concept quickly extended to epigenetic replication errors on DNA methylation (DNAm)[12,13]. During DNA replication, epigenetic covalent modifications are not faithfully replicated to the daughter strand, resulting in stochastic DNAm changes. Stochastic DNAm fluctuation has been applied to infer the mean mitotic count of cell population[14-16]. On population scale, irreversible, stochastic DNAm changes were thought to be underlying age-associated DNAm change. A vast number of literatures has documented age-associated DNAm change[17], including hypomethylation and hypermethylation, in specific genomic regions. We term such genomic regions

clock-like differential methylation loci (ClockDML) because DNA methylation on them exhibited timekeeper-like behavior.

DNAm-based regression model predicts the age of biological sample at extremely high precision in many organisms[18-24], and correlates with rejuvenation or accelerated aging in various scenarios[25]. Similar age association of DNA methylation was conserved across mammalian species on homologous genomic regions[26], suggesting that it is controlled by a defined, possibly functional molecular mechanism[27]. Interestingly, predictor model has been built to estimate mitotic age of sample from DNAm state of a defined set of CpG loci[28,29], indicating that mitosis is associated with clock-like DNA methylation changes on specific genomic loci. Introducing population statistics into the DNAm-based age prediction model enabled single cell age prediction from single cell methylation sequencing data[30,31] suggesting that clock-like DNAm changes are not merely a statistical phenomenon on population scale but also holds on single cell level.

Based on the intrinsic link between chromatin accessibility and DNA methylation[32-37], we hypothesized that age-dependent DNA methylation could be either resulted from, or result in, chromatin accessibility changes on clock-like differential methylated loci. If so, the derived mitotic age of single cells from scATAC data would serve as a powerful tool to delineate developmental trajectory. In theory, mitotic age is a ‘timekeeper’ tracker: the mitotic age of an ancestor cell is lower compared to its progeny, and cells originated earlier in time should show lower mitotic age compared to those originated latter. Such measure of cell age, if exist, would provide a precise temporal reference of cell birth sequence to help delineating the developmental trajectories in a complex organism.

To develop a mitotic age estimator for scATAC data, we determined ClockDML across human genome and characterized the chromatin accessibility changes on these loci associated with cell mitotic age. The heterogeneity of chromatin accessibility on these loci reduced across cell division. We leveraged this phenomenon to develop a computational framework, *EpiTrace*, which infers cell mitotic age from scATAC data by counting the opened fraction of ClockDML in single cells.

Results

Chromatin accessibility on ClockDML enable cell age estimation

Although the molecular mechanism which generates age-associated DNAm change is unclear, it is possible that the methylation state of ClockDML might be affected by their chromatin accessibility. Alternatively, methylation state of ClockDML might reversely regulate chromatin accessibility. In either case, the chromatin accessibility on ClockDML could be used to deduce cell age (Figure 1a). However, the dynamics of chromatin accessibility on ClockDML during aging is currently unknown.

Literature documented ClockDML[18,25,26,28] were found mainly with methylation-specific microarrays and thus only represents a tiny fraction of possible genome-wide DNAm variation during aging. However, ATAC -- and more so scATAC -- data are too sparse to fully cover these loci. To enable accurate tracking of cell mitotic age by ATAC signal, we determined a large number of ClockDML in human genome by bisulfite capture sequencing the CpG island regions of peripheral blood mononuclear cells from a panel of donors of different ages, and determined the correlation coefficient between age and the methylation level (beta) on each loci (Extended Data Table 1: ClockDML). DNAm status of these ClockDML show excellent correlation with age in the training cohort (Figure 1b). Both general linear model (GLM) and probability model (TIMEseq)[30] built upon beta values of our ClockDML predicts donor age at good precision in an additional validation cohort of samples (R^2=0.8553 (GLM) / 0.7998 (TIMEseq), Extended Data Figure 1), indicating DNAm status of these loci stably drifts over age.

Functional annotation suggests that ClockDML are enriched in the open, accessible chromatin region of genome across different cell types and organs[38-41] (Figure 1c). In correlation of DNAm status, the fraction of opened ClockDML (hereinafter in short:

ClockAcc) shifts in correlation to cell aging (Supplementary Notes and Extended Data Figure 2-5). We established an algorithm, *EpiTrace*, that predicts sample age via counting the fraction of opened ClockDML in bulk ATAC-seq datasets (Supplementary Notes). Validation experiments using bulk ATAC-seq of FACS-sorted blood cells, iPSC induction experiments, and native immune cells showed that *EpiTrace* accurately predicted sample age in concordance of known developmental trajectories (Supplementary Notes). We then adopted the algorithm to single cell ATAC-seq data. In brief, ClockAcc is measured for each cell. Reduction of dispersion was performed by clustering cells via correlation of top variable ATAC peaks to form a cell-cell similarity matrix for diffusion-regression iterations of ClockAcc. The regularized ClockAcc were subjected to rank-based re-normalization to estimate cell age. To overcome sampling sparseness in scATAC-seq, we reasoned that it might be unnecessary for age-dependent chromatin accessibility to be always accompanied by DNAm change (Supplementary Notes). Thus, we perform stepwise iterations by extracting additional open regions with high correlation coefficient with estimated single cell age, then include them together with ClockDML to form a new set of clock-like loci for next round of analysis, until the age prediction converged (Figure 1d).

As a proof-of-concept, we gathered single cell ATAC data from various studies of early human embryonic development from gametes to blastula[42, 43] (Figure 1e), and subjected them to *EpiTrace* analysis without batch correction (Figure 1f). Single cell total ClockAcc positively correlates with known cell mitotic age. Whilst the initial *EpiTrace age* prediction is noisy, iterative optimization improved signal-to-noise ratio to draw a biologically plausible trajectory of age resetting during early embryonic development: starting from zygote, cell mitotic age gradually reduces to near ground

state at the time of zygotic genome activation at morula, before its re-bounce in inner cell mass (ICM) and trophectoderm (TE).

ClockAcc change occurs upstream of DNAm shift on ClockDML to report cell mitotic age

Both *EpiTrace* and DNAm on ClockDML predict biological age of the sample. To compare their performance, we took a single-cell multiome (scMulti) sequencing dataset[44] of primed human embryonic stem cell ('Primed' hESC) cultures undergoing chemical reprogramming through a '4CL naive PSC' state towards 8-cell-like ('8CL') state (Figure 2a), measured *EpiTrace* age in single cells, and compared the *EpiTrace* prediction with biological age predicted by whole-genome bisulfite sequencing (WGBS) of the same cultures[44]. Both DNAm-based prediction of sample age (Figure 2b) and single cell age predicted by *EpiTrace* (Figure 2c) suggest mitotic age increases as cells undergone transformation, with single cell age gradually increasing across the evolution trajectory towards 8CL state (Extended Data Figure 6). Furthermore, the biological age estimation of DNAm and *EpiTrace* are precisely correlated (correlation coefficient of mean single cell (sc)EpiTrace age x mean DNAm age: 0.998 (P=0.04); scEpiTrace age x mean DNAm age: 0.526 (P=1.9e-38)) (Figure 2d). Whilst RNA velocity projections on these cells showed erroneous evolution trajectories root at single cells of a differentiated state (Extended Data Figure 7), combining RNA velocity and *EpiTrace* age of the same cell results in more biological plausible evolution trajectories with the primed hESC as root of all other cells (Figure 2e). These results suggest that chromatin accessibility on ClockDML predicts single cell biological age as good as the DNAm-based age estimator.

The relationship between DNA methylation and chromatin accessibility on

ClockDML is elusive. We measured *EpiTrace* age in a scATAC dataset of cells undergoing the early stages of chemical reprogramming from differentiated endodermal cells towards induced pluripotent stem cells[45] (Figure 2f), and compared them with WGBS-predicted ages from the same study. *EpiTrace* age prediction of single cells significantly decreases at stage II compared to the uninduced state (Figure 2g). The decrease in chromatin accessibility-based age prediction is less affected by 5-azaC removal but strongly affected by JNK inhibitor removal. *EpiTrace* age shows negative correlation to peaks associated with genes acting in current and future stage, and positive correlation to peaks associated with genes acting in previous stages (Figure 2h). The age-peak correlation is robust as the same peak shows highly concordant correlation coefficients in independent experiments with different starting cell (Figure 2h). These results suggest relevance between the observed mitotic age resetting and cell fate reprogramming.

DNAm-predicted age of cells along the chemical induction procedure shows biological age first increase at induction stage II (Figure 2i) before decreasing to near zero at pluripotent state. Removal of 5-azaC from the induction formula blocks age increase at stage II, indicating the apparent DNAm age increase is a result of global DNA demethylation[30]. Comparing the DNAm age and *EpiTrace* age prediction of the same cell sets suggests that chromatin accessibility on ClockDML is epistatically upstream of DNAm change (Figure 2j). We conclude that chromatin accessibility changes correlate with cell mitotic age, and precedes the irreversible DNAm change on ClockDML. It is likely that DNAm changes on ClockDML is a consequence of chromatin accessibility shift.

***EpiTrace* in combination with mitochondrial mutation clock reveals**

age-dependent regeneration potential of HSC

To test *EpiTrace* age estimation in genetically defined cell lineages, we took advantage of a mitochondria-enhanced scATAC dataset[8] of cultured CD34 hematopoietic stem cells (HSC) undergone *in vitro* expansion for 14 days before being forced into differentiation under LIF/EPO towards myeloid/erythroid lineages for additional 6 days (Figure 3a). Cells were clustered by their transcriptomic (scRNA) phenotype as progenitors (Prog), differentiated (Diff) or terminally differentiated (Terminal) cells. Furthermore, they were segregated into lineages (clones) arise from the same progenitor by mitochondrial SNV.

EpiTrace age prediction is reversed for erythroid lineage, probably due to genome-wide chromatin condensation[46] (Extended Data Figure 4 and 8). For the myeloid lineage, *EpiTrace* age of cells from the same clone increased from progenitor to terminal myeloid cells (Figure 3b-c). Forced induction starting from Day 14 (D14) increased age of differentiated cells as expected but decreased age of terminal cells (Figure 3b-c). To explore this phenomenon in depth, we classified clones according to the relative age change between days of culture (Extended Data Figure 9): clones that exhibits age increase from D8 to D20 in a cluster as 'Aged', and those exhibit age decrease from D8 to D20 as 'Rejuvenated'. Whilst most progenitor and differentiated cells show clonal aging during the induction, clonal rejuvenation dominates the terminally differentiated clusters (Figure 3c-d and Extended Data Figure 9). The proportion of clones showing rejuvenation increase in terminal cells, in correlation with their differentiation state (Figure 3c-d and Extended Data Figure 9), suggesting that these terminally differentiated cells were derived from younger hematopoietic progenitor instead of existing intermediate differentiated cells (Figure 3e).

Single cell age estimation facilitates discovery of molecular marker of peripheral-influx T cells underlying anti-PD1 response

The CD34 dataset demonstrated above is based on an ideal *in vitro* scenario with cells at similar starting point. To test *EpiTrace* in a complex cell population in *in vivo* setting with possible influx, efflux and proliferation, we applied *EpiTrace* to a scATAC dataset comprising biopsies from basal cell carcinoma pre- and post-anti-PD1 treatment[47] (Figure 4a). Post anti-PD1 treatment, cytotoxic T cells with exhaustion markers (Tex) are significantly increased in anti-PD1 responders (R) but not non-responders (NR). More immature Tex cells presented in non-responders and this phenomenon is exaggerated post anti-PD1 treatment. However, overall maturity did not change in responders. Importantly, *EpiTrace* age of interim and mature Tex in responders did not change post-anti-PD1, suggesting that the increased cell number might not be due to local proliferation of pre-anti-PD1 mature Tex (Figure 4b).

New post-anti-PD1 mature Tex cells could be either derived from pre-anti-PD1 immature Tex, or from influx of peripheral T cells. To test these alternatives, we performed correlation of chromatin accessibility on Tex differentially expressed peaks and cell age. Hierarchical clustering of the peak openness of pseudobulk cells from similar age and phenotype segregated peaks into three clusters: C1: response-specific, age-independent; C2: response-irrelevant and age-associated; and C3: non-response-specific peaks that weakly associated with age (Figure 4c-d). Interestingly, known markers of activated (TIGIT, LAYN, HAVCR2) and tumor-reactive T cells (ENTPD1) were segregated into different clusters. T cell exhaustion markers (TOX2, ID2, MAFA) and tissue resident Trm marker (CXCR6) were found to be mainly associated with age. anti-PD1 response is not associated with cell age, but instead with C1 peak expression. On the contrary, cell age is associated

with C2/C3 peaks which are related to no response under anti-PD1, indicating that the anti-PD1 responsiveness is associated with influx of a pre-existing cell population instead of local proliferation of cells. Particularly, we noticed that the tumor draining lymph node attraction markers CCR7 and SELL (CD62L) are expressed only in aged Tex, suggesting that they are recent migrate-in effector memory cells from draining lymph node (Figure 4e). Hence, in concordance with recent results with single cell TCR clonal tracing and clinical observations[48, 49], our analysis suggest that peripheral influx of tumor-reactive T cells underly anti-PD1 response. Finally, we identified that cis-regulatory loci of a co-receptor and negative regulator of TGF-beta, CD109 (Figure 4f), and the nicotinic acetylcholine receptor CHRNA1 (Figure 4g), are specifically activated in response-associated Tex cells, suggesting novel targets for future research.

EpiTrace reveals the developmental history during human cortical gyrification

To test how mitotic age estimation might complement RNA-based development analysis, we applied *EpiTrace* to a scMulti dataset from pcw21 human fetal brain cortex[50] to study the trajectory of glutaminergic neuron development (Figure 5a). Glutaminergic neurons develop from radial glia (RG), through the cycling progenitor (Cyc.Prog) cells into neuronal intermediate progenitor cells (nIPC), before it undergone a cascade of maturation (GluN1->2->3->4->5)[50-52]. We modeled the cell fate transition by CellRank[53] with kernels built with RNA velocity, CytoTrace, *EpiTrace* age, or a combined kernel with all three estimators. Whilst RNA velocity and CytoTrace produced inconsistent transition trajectory that points towards a group of nIPC (Figure 5b, i-ii), kernel with *EpiTrace* age revealed correct direction-of-development from nIPC towards terminally differentiated neuron (Figure

5b, iii). The combined kernel of all three estimators resulted in biologically plausible transition trajectory which starts from RG to bifurcate into two different branches, each give rise to a distinct nIPC population that differentiates into mature neurons (Figure 5a).

Two transcription factors, TCF4 and NR2F1 (COUP-TFI), are differentially expressed between the branches. They exhibit significant differential binding activities in these neurons (Figure 5c). Interestingly, NR2F1 is mainly expressed in the gyrus of human cortex, and hereditary NR2F1 loss-of-function mutation is associated with mental retardation and poly-microgyri phenotype[54-56]. NR2F1 TFBS-associated peaks are open in a branch (Figure 5d) which is NR2F1 negative (Figure 5e) and LMO3 positive (Figure 5f), suggesting that NR2F1 turned into a transcriptional repressor in nIPC. *EpiTrace* age of the $NR2F1^+$ branch nIPC is significantly higher compared to LMO3+ nIPC, suggesting accelerated mitotic activity (Figure 5g). In concordance with this, CytoTRACE score of $LMO3^+$ nIPC is higher than $NR2F1^+$ nIPC (Figure 5h). These results indicate that nIPC are divided into $NR2F1^+$ clones that support earlier neurogenesis and $LMO3^+/NR2F1^-$ clones that expand relatively later, linking the gyrus-specific expression pattern of NR2F1 to its function in cortical gyrification[54].

We compared EpiTrace age of the neurons with their CytoTRACE score (Figure 5i). Whilst CytoTRACE score of glutaminergic neurons correlate with their differentiation, the *EpiTrace* age of these cells are inversely correlated with their maturity. To explain this inconsistency, we built ‘phylogenetic tree’ with ClockDML chromatin accessibility (*EpiTrace* phylogeny). We reasoned that cells traverse on the phenotype manifold on branched trajectories whilst they undergone mitosis. As they evolve, chromatin accessibility on ClockDML converges into specific state which should be

lineage-dependent because of the irreversible nature of such change. Hence, it is possible to infer cell lineage tree using phylogenetic-like methods. Such analysis revealed a birth sequence of glutaminergic neurons: GluN5 are firstly divided from RG, followed by GluN4, GluN2, GluN3 and GluN1/nIPC (Figure 5j), indicating that neurons that formed earlier undergone a longer post-mitotic maturation (Extended Data Figure 10). Together, this analysis demonstrated that *EpiTrace* age analysis complements RNA velocity and stemness prediction in characterizing complex organ development, indicated a long post-mitotic maturation of neurons, and revealed the molecular mechanism of NR2F1 controlling human nIPC proliferation during cortical gyrification (Figure 5k).

Single cell lineage tracing by mitotic age loci indicates de-differentiation driven by master transcription factor in translocation renal cell carcinoma

We have already demonstrated *EpiTrace* can track development using *developing* tissues. To test if *EpiTrace* can recover epigenomic changes during development from a single, terminally-developed, static snapshot from adult tissue, we applied *EpiTrace* to a scATAC dataset from adult human kidney[57](Figure 6a). Birth sequence of kidney cells by *EpiTrace* phylogeny analysis suggests an endothelial origin of kidney tubules and delineates cell-type-specific generation cascade during nephrogenesis (Figure 6b), with correlation to their spatial position (Extended Data Figure 11). Distribution of *EpiTrace* age for each cell type suggesting a distal-to-proximal genesis cascade of nephron tubules with a late expansion of proximal tubule (Figure 6c).

Correlation between *EpiTrace* age and peak openness showed clear segregation of peaks opened in progenitor or differentiated proximal tubule (PT) cells (Figure 6d). Importantly, such association is not guided by known cell type information, indicating

the power of *EpiTrace* in positioning single cells along their evolution trajectory. Interestingly, translocation renal cell carcinoma (TRCC) driver gene TFEB is specifically activated in progenitor cell and shows age-dependent decrease of activity. In contrast, all hereditary renal dysgenesis (CAKUT) genes FGF8, FGFR2, SLIT3, GDNF and NHS are associated with differentiated cell-specific, age-dependent increased peaks. These results suggest that CAKUT is linked to genes functioning in terminal proximal tubule cell fate determination and function, whilst TRCC oncogenesis is linked to the mis-expression of progenitor-specific transcription factor, possibly forcing de-differentiation of terminally differentiated proximal tubule cells into a stem-like state.

Early-on branching evolution of glioblastoma clonal evolution tracked by single cell mitotic age

Finally, we used a human glioblastoma scATAC dataset to study whether *EpiTrace* age analysis could work for cell evolution in oncogenesis[58] (Figure 7a). In this tumor, copy number variation (CNV) analysis showed that MDM4 amplification dominates the malignant clones, which additionally have either EGFR or PDGFRA amplification, resulting in increased chromatin accessibility around these genes (Figure 7b-e). With *EpiTrace*, we identified a pre-malignant cluster (7) which is younger than all malignant clones (4/6/5/0/3) but shows accelerated aging/mitosis count compared to the 'normal clones' (1/9) (Figure 7b, f), is with lower MDM4 amplification (Figure 7c), and is without either EGFR or PDGFRA amplification (Figure 7d-e).

Interestingly, some MDM4+ cells are with both EGFR and PDGFRA amplification (Extended Data Figure 12). *EpiTrace* age analysis revealed that the $MDM4^+$ only cells are ancestral to triple-positive, $EGFR^+/PDGFRA^+$ cells, followed by loss of either

EGFR or PDGFRA in the progeny (Figure 7f). This is further supported by *EpiTrace* phylogeny analysis (Figure 7g). Branched evolution of $MDM4^+/EGFR^+$ and $MDM4^+/PDGFRA^+$ cells initiated at the beginning of malignancy transformation (Extended Data Figure 12). Together, these results characterized the evolution trajectory of malignancy from the $MDM4^+$ pre-malignant clone to the earliest malignant cell population with amplification on MDM4, PDGFRA and EGFR in a catastrophic genomic instability event, which bifurcates into heterogeneous clones with either PDGFRA or EGFR addiction (Figure 7h). *EpiTrace* age analysis revealed the pre-malignant state of this tumor, and suggested branching evolution of this tumor to indicate that heterogeneous cancer clones arise early in malignancy transformation. Although we sampled only a fraction of the tumor, the similar cell age estimated for $MDM4^+/EGFR^+$ and $MDM4^+/PDGFRA^+$ clones suggested that neither of these clones gained selective advantage during tumor growth. Instead, they are under neutral evolution.

Discussion

We formulated a model of clock-like chromatin accessibility change during cell mitotic aging, leading to the discovery of a universal epigenomic hallmark during cellular development: chromatin accessibility across ClockDML. The heterogeneity of chromatin accessibility across ClockDML reduces at each mitosis, resulting in converged, homogeneous activity pattern. We showed that chromatin accessibility changes act upstream of clock-like DNA methylation changes. Counting the fraction of opened ClockDML of each cell gives a simple, phenotypic measure of cell mitotic age. We leveraged such measure to build a tool, *EpiTrace*, to predict cellular mitotic age. Furthermore, we showed the similarity across ClockDML chromatin accessibility between single cell clusters can serve as an accurate distance measure for phylogenetic analysis.

DNA methylation shift on ClockDML is widely accepted as a hallmark of aging. However, the molecular mechanism generating age-dependent DNA methylation is yet unknown. Our data indicated that sample age predicted by chromatin accessibility and DNA methylation were significantly correlated, suggesting they are either under control of a similar biological process, or mutually controlling each other. With the chemical-induced iPSC data, we showed that chromatin accessibility on ClockDML changes independently from DNA methylation. These results suggest that age-dependent chromatin accessibility is likely to be epistatically upstream of age-dependent DNA methylation. Whether age-dependent DNA methylation is secondary to chromatin accessibility changes, and whether DNAm changes would in turn lock the chromatin accessibility state during aging, await future investigation.

It is well known that chromatin accessibility changes accompanied development and cell fate transition. However, we noticed that chromatin accessibility on ClockDML is phenotypic neutral – that is, irrelevant to cell phenotype – based on several evidence. Firstly, *EpiTrace* age measured on the same set of ClockDML works for different lineages, in various test cases. Secondly, as we have shown in examples such as cortical development and CD34 *in vitro* differentiation, *EpiTrace* age of single cell correlates with its accumulative mitosis number instead of developmental maturity. Thirdly, as we have shown in the cortical development, kidney, embryonic development, iPSC induction and chemical-induced 8-cell-like state examples, *EpiTrace* age does not correlate with stemness – or developmental potential – of the cell. The molecular mechanism generating such chromatin accessibility shift on ClockDML following mitosis is still unknown. We observed weak-but-widespread and heterogeneous chromatin accessibility across genome in progenitor or stem cells. Such phenomenon is in correlation with the observation of scRNA data, where expressed transcript number is correlated to stemness[3].

It is unexpected for the authors that the phylogenetic tree built with chromatin accessibility on ClockDML (*EpiTrace* phylogeny) for single cell clusters of the same developmental lineage is highly accurate. In fact, we noticed that such phylogenetic trees outperform those built with the highly variable peaks from scATAC data, in terms of accuracy. Despite the fact that ClockDML are highly enriched in cis-regulatory regions, there is no functional enrichment of them in specific developmental pathways, or specific types of genomic elements (besides active cis-regulatory elements). Furthermore, this phenomenon is also phenotypic neutral. These results not only suggest a consistent birth sequence of cell types within the

lineage, but also indicates that senescence is a defined molecular process across cell types.

The current study is not without limitation. We noticed that inferred mitotic age by current *EpiTrace* algorithm is dependent on sequencing depth: age inferred from sparse sequencing is reversed from that inferred from saturated sequencing, and cell lineages that differentiate with massive global genome condensation such as erythroid[59] might show reversed mitotic age compared to those without. This phenomenon is related to single cell population heterogeneity across development, the nature of chromatin accessibility shift during mitotic aging, the statistical model underlying our algorithm, and the limitations of sequencing technique. Future improvement of the algorithm might require in depth study of molecular mechanisms driving age-dependent chromatin accessibility changes on these loci. We also noted an apparent limitation of the study is that all the validation are based on limited human scATAC datasets. Such limitation is due to a lack of genome-wide ClockDML reference dataset for other animals. We expect future efforts would resolve this and extend the application range of our algorithm.

In conclusion, we discovered mitosis-associated, age-dependent chromatin accessibility on ClockDML, and developed computational methods to track single cell age with it. By comparison studies, we showed that the chromatin accessibility-based mitosis age measure complements somatic mutation, RNA velocity and stemness predictions, to predict cell evolution trajectory with high precision and power. We expect *EpiTrace* would be a useful tool for single cell studies for delineation of cellular hierarchies and organismal aging.

Methods

Human biospecimens

This study was conducted in accordance with the measures of the Declaration of Helsinki and the Ethic Protocols of Human Genetic Resource Preservation Center of Hubei Province, China (Hubei Biobank). Blood samples from healthy adult donors (total n=71 including: training n=24; validation n=47) used in this study were pretreated and preserved by the Hubei Biobank, approved by the Institutional Ethical Review Board (approval number: 2017038). Informed consents were obtained from all individuals.

Single-stranded bisulfite capture sequencing

Genome-wide CpG bisulfite capture sequencing was performed as in Xiao *et al.*[40] and Yin *et al.*[60] using the Tequila-7N protocol. In brief, genomic DNA were extracted from PBMC, bisulfite-converted, poly-dT-tailed with terminal transferase, ligated with a poly-dA-extruded 3' adaptor, linearly amplified for 12 cycles with a primer that complements with the 3' adaptor, and ligated with a 5' adaptor with 7 N-nucleotide overhung. The resulted adaptor-ligated single-stranded inserts were amplified with Illumina P5/P7 sequencing adaptor primers and captured with EpiGiant probe set (Nimblegen, Cat. #07138881001) following the manufacturer's protocol. Post-capture libraries were retrieved, amplified, quantified, and sequenced on Illumina Novaseq with PE150 format to 50M reads.

Pre-processing of sequencing data

The bisulfite capture sequencing (methylation) data were processed exactly as described in Yin *et al.*[60]. Per-CpG DNA methylation frequency (beta) was computed with PileOMeth (v0.1.13-3-gca82747, https://github.com/bgruening/PileOMeth). Bulk ATAC and Cut-And-Run sequencing data were processed exactly as described in Yin *et al.*[60], with hg19 reference genome. 10× scATAC sequencing data were processed by CellRanger-ATAC (v1.2.0) with hg38 reference genome. 10x scMultiomic sequencing data were processed by CellRanger-ARC (v2.0.1) with hg38 reference genome. Fragment files from these pre-processed ATAC data were combined and analyzed in ArchR (v1.0.1)[6]. Quality control, doublet removal, dimensionality reduction (LSI), clustering and peak finding were all performed in ArchR according to its reference. Annotation of cells from their original publications/dataset were taken without adjustment. Cells were not manually excluded from the analysis, except in three cases: 1. Aneuploid 3PN cells in human embryonic development dataset; 2. PT3 cluster in kidney scATAC dataset, which is donor-specific; 3. Erythroid lineage in *in vitro* CD34 differentiation dataset, of which the *EpiTrace* age estimation is aberrant due to genome-wide condensation of chromatin -- and is discussed in the main and supplementary text. Lift-over between reference genome sets were executed with easyLift (v0.2.1, https://github.com/caleblareau/easyLift). 10x scRNA data were aligned to hg38 reference transcriptome by kallisto (v0.46.1, https://github.com/pachterlab/kallisto) and converted to splice/unsplice count matrix, which are passed to Seurat[61] (v4) for

basic quality control, normalization, scaling, dimensionality reduction (PCA), and clustering. Seurat object was converted to anndata format with SeuratObjects (v4.0.4, https://github.com/mojaveazure/seurat-object) prior to RNA velocity and CytoTRACE analysis.

Age-dependent clock-like DML

Age-dependent, clock-like DML (Chronology) were discovered in the training cohort, with correlating per-CpG beta value against sample donor age using an adapted code from scAge (Trapp *et al.* 2021[30], and personal communication with A.Trapp). Cut-off of the absolute value of Pearson's Correlation between sample age and beta was set to 0.7. Known clock-like mitosis-associated DML from Yang *et al.*[28] (Mitosis) and development-associated DML from Zhou *et al.*[38] (solo-WCGW) were also included for comparison. Enrichment of clock-like DML on cis-regulatory elements from human single cell ATAC[39], hematopoiesis cell ATAC[62], bladder cancer single cell ATAC[40], placenta single cell ATAC (Gong *et al.*, unpublished) datasets were performed with Fisher's exact test in R. Clock-like DML are clustered according to their beta and sample age using *hclust* in R (4.1.3) and unsupervised classification was performed by *cutree*. Mean beta of each group of Clock-like DML were correlated with age. The group (G8) of clock-like DML with best negative correlation between mean beta and sample age was shown in Figure 1b.

Predicting sample age by DNAm

Age prediction by DNAm beta values were performed by either GLM model trained according to Horvath *et al.*[25] or a modified scAge code (TIMEseq model)[30,63]. The training cohort did not include donor aged 0-18 years thus no log-transformation was performed to the younger age. For GLM model, the loci were filtered for sufficient coverage >30x. For TIMEseq model, the loci were not filtered.

Measuring the diversity of chromatin accessibility on ClockDML region

ATAC peaks overlapping ClockDML were pulled with GenomicRanges *findOverlaps* and peak-wise counts, inter-sample peak mean, median, standard deviation and coefficient-of-variation (CV) were calculated in R (v4.1.3) package sparseMatrixStats (v1.6.0). The entropy of ClockDML chromatin accessibility was measured by firstly normalizing per-peak read counts to total counts over all peaks of interest to have a vector of probability of observing a read in peak x (p(x)), and compute the Shannon entropy as: Entropy = $-1\times\text{sum}(p(x)\times\log(p(x))$.

Measuring general accessibility of ClockDML region

ATAC peaks overlapping ClockDML were pulled with GenomicRanges *findOverlaps* and peak-wise counts were censor-normalized according to Qiu *et al.*[64]. ClockAcc, defined as the number of total opened DML, were measured for each cell for peaks with nonzero read coverage. To perform cell-cell-similarity-based reaction-diffusion smoothing, cell-cell similarity was firstly calculated: Inter-single-cell dispersion of

each peak was calculated, and used to select the top 5% variable peaks. Cell-cell similarity (correlation) matrix was computed as correlation of top 5% variable peaks between each single cell using WGCNA::cor (v1.70-3). The correlation matrix was further thresholded by its mean, with correlation values lower than mean being set to zero. Finally, the correlation vector for each cell was normalized by dividing the total sum of correlation.

Assuming the mitotic age acts on the phenotype of cell (manifested as a cell-cell-similarity matrix) to result in smoothened ClockAcc, the mitotic age of cells could be solved by non-negative least square. Non-negative least square decomposition of the cell-cell-similarity matrix and smoothened ClockAcc was computed by Lawson-Hanson algorithm using nnls::nnls (v1.4). Reaction-diffusion regression of the NNLS result was performed, using an HMM-like approach: in each iteration, the NNLS result is updated to the weighted sum of current NNLS result and the cross product of NNLS result with cell-cell similarity matrix, until the difference between updated NNLS result and current NNLS result is lower than preset threshold. The cross product of final NNLS result with cell-cell similarity matrix results in a smoothened, regressed measure of ClockAcc. The overall algorithm was first proposed in CytoTRACE[3]. The resulting smoothed total chromatin accessibility were re-normalized between 0 to 1 to give a rank of samples. For bulk sequencing, the rank is reversed as (1-rank). Such rank denotes the relative birthday (mitotic age) of the cell within the population.

Algorithm to measure general accessibility of ClockDML region in single cell ATAC data

The algorithm is similar to bulk ATAC chromatin accessibility measurement on ClockDML, with additional iterations by inclusion of additional loci-of-interest. Briefly, after the re-normalization of smoothed total chromatin accessibility, the rank is not reversed. For iteration, read counts of all ATAC peak were correlated against single cell age estimation (the rank). The correlation coefficient is scaled and peaks with a sufficiently high correlation coefficient vs age (normally Z>2.5 or 3) was included in the new 'Clock-like loci' set together with all known ClockDML. In the next round of iteration, the new 'Clock-like loci' set replaces ClockDML for computation. The iteration was performed for designated times or until convergence of age estimation. Final single cell *EpiTrace* age estimation was given by the rank of smoothed total chromatin accessibility over clock-like loci in final iteration. No reversal of ranking was performed for single cell datasets.

Phylogenetic tree construction with chromatin accessibility on ClockDML region

Mitosis birth sequence-based phylogenetic tree was inferred with chromatin accessibility on ClockDML in Seurat with the function BuildClusterTree and rooted using package ape (v5.6-1) with the cell cluster which has minimal mean *EpiTrace* age serve as outgroup.

Association of chromatin accessibility to estimated cell age

Read counts of all ATAC peak were correlated against sample age estimation with R (v4.1.3) package WGCNA and BiocParallel (v1.28.3). Correlation coeffcients were scaled by R. Age-dependent chromatin accessibility trend of loci with significant correlation were analyzed with tradeSeq (v1.8.0). To generate heatmap for visualization, chromatin accessibility of the same locus is averaged for cells of similar age bin (0-100) and phenotype.

Mitochondrial SNV (mGATK) analysis

mGATK analyses were performed following Lareau *et al.*[8] to reproduce the clones described in the original publication.

RNA velocity and CytoTRACE analysis

Spliced and unspliced scRNA data were exported from R (v4.1.3) with SeuratObjects to h5ad format and imported into python (v3.9.7) to be processed by CellRank (v1.5.1, https://github.com/theislab/cellrank)[53]. The scVelo and CellRank CytoTRACE kernel were used for estimating RNA velocity and CytoTRACE score, respectively. scVelo (v0.2.4, https://github.com/theislab/scvelo)[65] was used to produce velocity embeddings upon UMAP space.

Multiomic derivation of cell evolution trajectory with kernel combining cell age and RNA velocity

A custom kernel of CellRank built upon *EpiTrace* age was used to project *EpiTrace* age into CellRank. A combined kernel of RNA velocity and *EpiTrace* was built with 0.75×velocity + 0.25×age, which was used in the downstream GPCCA estimator in CellRank. Lineages were calculated by GPCCA, with cell clusters with least age were designated as root and cell clusters with maximal age were designated as terminal. Velocity genes were computed by the GPCCA estimator for given lineage.

Transcription factor motif differential analysis

TFBS motif differential activity were determined by Signac (v1.5.0, https://github.com/timoast/signac)[66] using the same *EpiTrace* object built during *EpiTrace* analysis.

Copy number variation analysis

scATAC copy number variation were called with CopySCAT (v0.3.0, https://github.com/spcdot/CopyscAT)[58] and annotated with ChIPseeker (v1.30.3)[67].

References

1. Buenrostro, J.D. et al. Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* **523**, 486-490 (2015).
2. La Manno, G. et al. RNA velocity of single cells. *Nature* **560**, 494-498 (2018).
3. Gulati, G.S. et al. Single-cell transcriptional diversity is a hallmark of developmental potential. *Science* **367**, 405-411 (2020).
4. Teschendorff, A.E. & Enver, T. Single-cell entropy for accurate estimation of differentiation potency from a cell's transcriptome. *Nat Commun* **8**, 15599 (2017).
5. Erhard, F. et al. scSLAM-seq reveals core features of transcription dynamics in single cells. *Nature* **571**, 419-423 (2019).
6. Granja, J.M. et al. ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nat Genet* **53**, 403-411 (2021).
7. Ludwig, L.S. et al. Lineage Tracing in Humans Enabled by Mitochondrial Mutations and Single-Cell Genomics. *Cell* **176**, 1325-1339 e1322 (2019).
8. Lareau, C.A. et al. Massively parallel single-cell mitochondrial DNA genotyping and chromatin profiling. *Nat Biotechnol* **39**, 451-461 (2021).
9. Xu, J. et al. Single-cell lineage tracing by endogenous mutations enriched in transposase accessible mitochondrial DNA. *Elife* **8** (2019).
10. Vaziri, H. et al. Evidence for a mitotic clock in human hematopoietic stem cells: loss of telomeric DNA with age. *Proc Natl Acad Sci U S A* **91**, 9857-9860 (1994).
11. Hills, M., Lucke, K., Chavez, E.A., Eaves, C.J. & Lansdorp, P.M. Probing the mitotic history and developmental stage of hematopoietic cells using single telomere length analysis (STELA). *Blood* **113**, 5765-5775 (2009).
12. Yatabe, Y., Tavare, S. & Shibata, D. Investigating stem cells in human colon by using methylation patterns. *Proc Natl Acad Sci U S A* **98**, 10839-10844 (2001).
13. Kim, J.Y., Tavare, S. & Shibata, D. Human hair genealogies and stem cell latency. *BMC Biol* **4**, 2 (2006).
14. Lopez-Garcia, C., Klein, A.M., Simons, B.D. & Winton, D.J. Intestinal stem cell replacement follows a pattern of neutral drift. *Science* **330**, 822-825 (2010).
15. Snippert, H.J. et al. Intestinal crypt homeostasis results from neutral competition between symmetrically dividing Lgr5 stem cells. *Cell* **143**, 134-144 (2010).
16. Gabbutt, C. et al. Fluctuating methylation clocks for cell lineage tracing at high temporal resolution in human tissues. *Nat Biotechnol* (2022).
17. Bell, C.G. et al. DNA methylation aging clocks: challenges and recommendations. *Genome Biol* **20**, 249 (2019).
18. Bocklandt, S. et al. Epigenetic predictor of age. *PLoS One* **6**, e14821 (2011).
19. Wang, M. & Lemos, B. Ribosomal DNA harbors an evolutionarily conserved clock of biological aging. *Genome Res* **29**, 325-333 (2019).
20. Meer, M.V., Podolskiy, D.I., Tyshkovskiy, A. & Gladyshev, V.N. A whole lifespan mouse multi-tissue DNA methylation clock. *Elife* **7** (2018).
21. Stubbs, T.M. et al. Multi-tissue DNA methylation age predictor in mouse. *Genome Biol* **18**, 68 (2017).
22. Thompson, M.J. et al. A multi-tissue full lifespan epigenetic clock for mice. *Aging (Albany NY)* **10**, 2832-2854 (2018).
23. Petkovich, D.A. et al. Using DNA Methylation Profiling to Evaluate Biological Age and Longevity Interventions. *Cell Metab* **25**, 954-960 e956 (2017).
24. Mayne, B. et al. A DNA methylation age predictor for zebrafish. *Aging (Albany NY)* **12**, 24817-24835 (2020).
25. Horvath, S. DNA methylation age of human tissues and cell types. *Genome Biol* **14**, R115 (2013).
26. Mammalian Methylation, C. et al. Universal DNA methylation age across mammalian tissues. *bioRxiv*, 2021.2001.2018.426733 (2021).

27. Williams, G.C. Pleiotropy, Natural-Selection, and the Evolution of Senescence. *Evolution* **11**, 398-411 (1957).
28. Yang, Z. et al. Correlation of an epigenetic mitotic clock with cancer risk. *Genome Biol* **17**, 205 (2016).
29. Youn, A. & Wang, S. The MiAge Calculator: a DNA methylation-based mitotic age calculator of human tissue types. *Epigenetics* **13**, 192-206 (2018).
30. Trapp, A., Kerepesi, C. & Gladyshev, V.N. Profiling epigenetic age in single cells. *Nature Aging* **1**, 1189-1201 (2021).
31. Kerepesi, C., Zhang, B., Lee, S.G., Trapp, A. & Gladyshev, V.N. Epigenetic clocks reveal a rejuvenation event during embryogenesis followed by aging. *Sci Adv* **7** (2021).
32. Klemm, S.L., Shipony, Z. & Greenleaf, W.J. Chromatin accessibility and the regulatory epigenome. *Nat Rev Genet* **20**, 207-220 (2019).
33. Wiench, M. et al. DNA methylation status predicts cell type-specific enhancer activity. *EMBO J* **30**, 3028-3039 (2011).
34. Pott, S. Simultaneous measurement of chromatin accessibility, DNA methylation, and nucleosome phasing in single cells. *Elife* **6** (2017).
35. Lee, H.J. et al. Regenerating zebrafish fin epigenome is characterized by stable lineage-specific DNA methylation and dynamic chromatin accessibility. *Genome Biol* **21**, 52 (2020).
36. Pandiyan, K. et al. Functional DNA demethylation is accompanied by chromatin accessibility. *Nucleic Acids Res* **41**, 3973-3985 (2013).
37. Thurman, R.E. et al. The accessible chromatin landscape of the human genome. *Nature* **489**, 75-82 (2012).
38. Zhou, W. et al. DNA methylation loss in late-replicating domains is linked to mitotic cell division. *Nat Genet* **50**, 591-602 (2018).
39. Zhang, K. et al. A single-cell atlas of chromatin accessibility in the human genome. *Cell* **184**, 5985-6001 e5919 (2021).
40. Xiao, Y. et al. Emergence of an adaptive epigenetic cell state in human bladder urothelial carcinoma evolution. *bioRxiv*, 2021.2010.2030.466556 (2021).
41. Corces, M.R. et al. The chromatin accessibility landscape of primary human cancers. *Science* **362** (2018).
42. Liu, L. et al. An integrated chromatin accessibility and transcriptome landscape of human pre-implantation embryos. *Nat Commun* **10**, 364 (2019).
43. Wu, J. et al. Chromatin analysis in human early development reveals epigenetic transition during ZGA. *Nature* **557**, 256-260 (2018).
44. Mazid, M.A. et al. Rolling back human pluripotent stem cells to an eight-cell embryo-like stage. *Nature* (2022).
45. Guan, J. et al. Chemical reprogramming of human somatic cells to pluripotent stem cells. *Nature* (2022).
46. Zhao, B. et al. Nuclear Condensation during Mouse Erythropoiesis Requires Caspase-3-Mediated Nuclear Opening. *Dev Cell* **36**, 498-510 (2016).
47. Satpathy, A.T. et al. Massively parallel single-cell chromatin landscapes of human immune cell development and intratumoral T cell exhaustion. *Nat Biotechnol* **37**, 925-936 (2019).
48. Liu, B. et al. Temporal single-cell tracing reveals clonal revival and expansion of precursor exhausted T cells during anti-PD-1 therapy in lung cancer. *Nat Cancer* **3**, 108-121 (2022).
49. Li, T. et al. T Cells Expanded from PD-1(+) Peripheral Blood Lymphocytes Share More Clones with Paired Tumor-Infiltrating Lymphocytes. *Cancer Res* **81**, 2184-2194 (2021).
50. Trevino, A.E. et al. Chromatin and gene-regulatory dynamics of the developing human cerebral cortex at single-cell resolution. *Cell* **184**, 5053-5069 e5023 (2021).
51. Llinares-Benadero, C. & Borrell, V. Deconstructing cortical folding: genetic, cellular

and mechanical determinants. *Nat Rev Neurosci* **20**, 161-176 (2019).

52. Kriegstein, A. & Alvarez-Buylla, A. The glial nature of embryonic and adult neural stem cells. *Annu Rev Neurosci* **32**, 149-184 (2009).

53. Lange, M. et al. CellRank for directed single-cell fate mapping. *Nat Methods* **19**, 159-170 (2022).

54. Bertacchi, M. et al. NR2F1 regulates regional progenitor dynamics in the mouse neocortex and cortical gyrification in BBSOAS patients. *EMBO J* **39**, e104163 (2020).

55. Bosch, D.G. et al. NR2F1 mutations cause optic atrophy with intellectual disability. *Am J Hum Genet* **94**, 303-309 (2014).

56. Naka, H., Nakamura, S., Shimazaki, T. & Okano, H. Requirement for COUP-TFI and II in the temporal specification of neural stem cells in CNS development. *Nat Neurosci* **11**, 1014-1023 (2008).

57. Wang, Q. et al. Single-cell chromatin accessibility landscape in kidney identifies additional cell-of-origin in heterogenous papillary renal cell carcinoma. *Nat Commun* **13**, 31 (2022).

58. Nikolic, A. et al. Copy-scAT: Deconvoluting single-cell chromatin accessibility of genetic subclones in cancer. *Sci Adv* **7**, eabg6045 (2021).

59. Ryzhkova, A. & Battulin, N. Genome Reorganization during Erythroid Differentiation. *Genes (Basel)* **12** (2021).

60. He, W. et al. Developmentally Delayed Epigenetic Reprogramming Underlying the Pathogenesis of Preeclampsia. *bioRxiv*, 2020.2005.2008.085290 (2020).

61. Stuart, T. et al. Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902 e1821 (2019).

62. Corces, M.R. et al. Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. *Nat Genet* **48**, 1193-1203 (2016).

63. Griffin, P.T. et al. Ultra-cheap and scalable epigenetic age predictions with TIME-Seq. *bioRxiv*, 2021.2010.2025.465725 (2021).

64. Qiu, X. et al. Single-cell mRNA quantification and differential analysis with Census. *Nat Methods* **14**, 309-315 (2017).

65. Bergen, V., Lange, M., Peidli, S., Wolf, F.A. & Theis, F.J. Generalizing RNA velocity to transient cell states through dynamical modeling. *Nat Biotechnol* **38**, 1408-1414 (2020).

66. Stuart, T., Srivastava, A., Madad, S., Lareau, C.A. & Satija, R. Single-cell chromatin state analysis with Signac. *Nat Methods* **18**, 1333-1341 (2021).

67. Yu, G., Wang, L.G. & He, Q.Y. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* **31**, 2382-2383 (2015).

Code and Data Availability

EpiTrace and accompanied source data is available at https://github.com/MagpiePKU/EpiTrace. Human pan-body scATAC data is downloaded from http://catlas.org/humanenhancer. TCGA pan-cancer ATAC-seq data is downloaded from https://gdc.cancer.gov/about-data/publications/ATACseq-AWG. Hematopoietic cell bulk and single-cell ATAC-seq data is from NCBI GEO GSE74310 and GSE96769. iPSC bulk ATAC data is from NCBI GEO GSE110544. T cell ATAC data is from NCBI GEO GSE179606. Human embryo development data is from NCBI GEO GSE101571 and CNGB CNP0000193. Chemical iPSC reprogramming data is from NCBI GEO GSE188461. Chemical induction of 8-cell-like state data is from https://figshare.com/s/ff707bf8242f7b3ed8f5, https://figshare.com/s/760d3ff54f1214a50cc2 and https://figshare.com/s/9c01c3b58d34b80de230. mtscATAC data is from https://github.com/caleblareau/mtscATACpaper_reproducibility. anti-PD1 treated cancer biopsy data is from https://github.com/GreenleafLab/MPAL-Single-Cell-2019. Cortical scMultiome data is from https://github.com/GreenleafLab/brainchromatin. Kidney scATAC data is from NCBI GEO GSE166547. Glioblastoma scATAC data is from NCBI GEO GSE139136 and GSE163656.

Author contributions

Conceive of idea: YZ, WJ, YX, XW; Algorithm development: YZ, WJ; Human biological specimen collection: XW, YX, LJ, GW, KQ; Data collection: FC; Sequencing experiment: WJ; Bioinformatic analysis: YZ, WJ; Biology and clinical analysis: YZ, YX, WJ; Writting the manuscript: YZ, WJ, YX, LJ, GW; Supervised the study: YZ, YX, XW.

Conflict of interests

All authors declare that they have no conflict of interest or financial conflicts to disclose.

Acknowledgments

The non-binary scAge DNAm-based age prediction code was kindly provided by Dr. Alex Trapp. The authors would like to thank all the authors sharing publicly available datasets used in this study. The authors would like to thank Drs. Yangyang Liu, Junqing Ye, Jie Fu, Ge Gao and Jianrong Yang for discussion. Part of the analysis was performed on the High Performance Computing Platform of the Center for Life Science (Peking University). This study is partially supported by grants from the National Natural Science Foundation of China (82172985, 81902603), Improvement Project for Theranostic Ability on Difficulty Miscellaneous Disease from National Health Commission of China (ZLYNXM202006), Science and Technology Department of Hubei Province Key Project (2022EJD001), Research Fund from Medical Sci-Tech Innovation Platform of Zhongnan Hospital of Wuhan University (PTXM2021023), Non-profit Central Research Institute Fund of Chinese Academy of Medical Sciences (2020-PT320-004) and Research Fund of Zhongnan Hospital of Wuhan University (ZNJC201920, ZNJC201915, SWYBK00-03, 413100049, KY0100000109). The funders played no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Figure legends

Figure 1. Chromatin accessibility change associated with irreversible DNA methylation drift on ClockDML enables cell age estimation. a, Schematic diagram of the underlying epigenetic mechanism of cell mitotic age tracing using chromatin accessibility (ChrAcc) on Clock-like differential methylated loci (ClockDML). **b,** DNA methylation level on ClockDML is linearly correlated with age to suggest irreversible age-dependent methylation change on these loci. **c,** Mitosis-associated and age (Chronology)-associated ClockDML, but not the solo-WCGW loci, are enriched on cis-regulatory elements with high chromatin accessibility in ATAC assay. OR: odds ratio of enrichment by Fisher's exact test. Mean and the estimate of upper/lower bound (whiskers) for OR are shown. **d,** Overview of *EpiTrace* algorithm. **e,** UMAP projection of human early embryonic development scATAC dataset. **f,** The total chromatin accessibility on ClockDML (ClockAcc), HMM-smoothened ClockAcc, initial and iterative *EpiTrace* ranking result (EpiTrace Age) corresponding to the embryonic dataset. HMM-smoothing and switching from raw chromosomal accessibility to relative ranks helps to segregate cell clusters, whilst iteration facilitates to include trajectory-specific clock-like loci for smoother, more biological plausible age prediction. Rejuvenation of cell age starts from zygote, continues across ZGA and finishes at morula, in correlation of genome-wide DNA demethylation. Abbrevations: ZGA: zygotic genome activation; ICM: inner cell mass; ESC: embryonic stem cell; TE: trophectoderm.

Figure 2. ClockAcc change occurs upstream of DNAm shift on ClockDML to report cell mitotic age. a, Schematic overview of the *in vitro* chemical induction of human pluripotent stem cell back to naïve (4CL) or 8-cell-like (8CL) states. **b,** Sample age estimated with DNA methylation of D0, D12 cultures and sorted 8-cell-like cells (8CLC) from D17 culture by Time-seq algorithm on WGBS data. **c,** Single cell age estimated with *EpiTrace* from the D17 scMultiome dataset. **d,** Correlation between single cell age estimated by *EpiTrace* and DNAm. **e,** Single cell evolution trajectories built with kernel combining *EpiTrace* age and RNA velocity information. **f,** Schematic overview of the *in vitro* chemical induction of adult fibroblasts towards human pluripotent stem cell. Both the uninduced and the intermediate stage II cultures are sequenced by scATAC. **g,** Single cell age estimated with *EpiTrace* from **(f)**. The induced cultures were either with full induction paradigm (+Chem: C6NYSA+T5J) or with 5-azaC/JNKin8 removed (-5aza, -JNKin8). Statistical comparisons were shown between groups (***: P<0.001; Wilcox test). **h,** ATAC peaks correlated with *EpiTrace* age estimated in MSC (x axis) or FB (y axis) cultures. Peaks of interest were labelled, colored by their corresponding developmental stage. Significant association of non-resetted age was found for peaks specific for previous stage (FB: Fibroblast; Development) and association of reseted age was found for peaks specific for StageII and pluripotency **(i)** Prediction of sample age by DNAm from WGBS data of chemical induction of iPSC (dataset GSE178966,

related to Figure 2f/g). Chemical reprogramming induces genome-wide demethylation and increase of DNAm age as reported in Trapp *et al.*[30], whilst addition of 5-azaC globally reduces DNA methylation to increase DNAm age as reported in Guan *et al.*[45]. Removal of 5-azaC blocks DNAm age from increasing. **j,** Scatter plot of WGBS DNAm age (x axis) and mean single cell *EpiTrace* age (y axis) of the same sample, suggesting chemical reprogramming with JNK inhibitor and H3K4 demethylation inhibitor directly regulate chromatin accessibility on ClockDML, whilst 5-azaC treatment directly regulate DNA methylation on ClockDML. This result indicates that DNA methylation changes occurs downstream of chromatin accessibility changes on ClockDML. Abbreviations: HEF: human embryonic fibroblast; hADSC: human adipose stromal cell (mesenchymal stromal cell); C6NYSA: combination of CHIR99021, 616452, TTNPB, Y27632, SAG, and ABT869; JNKIN8: c-Jun N terminal kinase inhibitor; 5-azaC: 5-azacytidine; T5J: tranylcypromine, JNKIN8 and 5-azaC.

Figure 3. *EpiTrace* in combination with mitochondrial mutation clock reveals age-dependent regeneration potential of HSC. a, Schematic overview of *in vitro* expansion-differentiation experiment. 800 Adult $CD34^+$ hematopoietic stem cells are cultured *in vitro* for 14 days before induced with LIF/EPO for differentiation. A fraction of D0/D8/D14/D20 cultures were subjected to mitochondria-enhanced scATAC sequencing. Prog: progenitor state; Diff: differentiated intermediate state; Terminal: terminal cell states. **b,** Examples (two) of mitochondrial SNV-derived cell clones during the time course of experiment. Cells are colored with *EpiTrace* age and projected to UMAP space. **c,** Per-clone mean *EpiTrace* age of the progenitor, differentiated and terminal populations over time. Age of the same clone are tracked (lines) across different time points. **d,** Percentage of clones exhibiting age decrease (blue) or increase (yellow) over time in each differentiation state. Abbreviations: prog: hematopoietic progenitor; prog_my: myeloid progenitor; my1-my4: intermediate-to-terminal differentiated myeloid cells. **e,** Schematic diagram summarizing the findings. During natural hematopoietic expansion, mitosis is associated with differentiation. However, under forced induction young HSC shows greater expansion potential such that most newly generated, terminally-differentiated cells are from them.

Figure 4. Single cell age estimation facilitates discovery of molecular marker of peripheral-influx T cells underlying anti-PD1 response. a, Schematic overview of experiment. Biopsies were taken from basal cell carcinoma patients before (pre-) and after (post-) anti-PD1 treatment and subjected to scATAC sequencing. **b,** Cell number (above) and *EpiTrace* age of 'exhausted T' (Tex) cells, separated by treatment response (R: responder; NR: non-responder) and phenotypic maturity (Immature/Interim/Mature). Significant increase in cell number but not cell age after PD1 treatment indicates peripheral influx of T cells into the tumor compartment. **c,** Heatmap showing scATAC peak activity in pseudobulk single cells grouped with phenotype (R/NR), sample (pre- or post-anti-PD1), and *EpiTrace* age. Correlation

between peak activity and *EpiTrace* age were shown on the left. Peaks were clustered according to their activity profile into response-specific, non-response-specific, and age-associated clusters. **d.** $CCR7^+$ $CD62L^+$ lymph-node residential cells are enriched in aged Tex in responders. **e,**ChrAcc (top) and cross-correlation between peaks (bottom) of CD109 loci from pseudobulk single cells grouped with phenotype (R/NR), sample (pre- or post-anti-PD1), and *EpiTrace* age (Young/Interim/Mature). Association of CD109 promoter ChrAcc across age is shown in the right panel. **f,** ChrAcc (top) and cross-correlation between peaks (bottom) of CHRNA1 loci from pseudobulk single cells grouped as in **(d)**. Association of CHRNA1 promoter ChrAcc across age is shown in the right panel. Both CD109 and CHRNA1 shows higher basal activity at young age in responders, suggesting these cells are 'primed' for PD-1 response.

Figure 5. *EpiTrace* reveals the developmental history during human cortical gyrification. a, UMAP projected cell evolution trajectory built with cellrank by using a hybrid kernel of *EpiTrace*, CytoTRACE, and RNA velocity, of a scMultiome-seq dataset from a pcw21 human brain. RG: radial glia; Cyc.Prog: cycling progenitor; nIPC: neuronal intermediate progenitor cell; GluN: excitatory glutaminergic neuron; IN: inhibitory GABAergic neuron; EC: endothelial cell; mGPC/OPC: medial ganglionic eminance progenitor/oligodendrocyte precursor cell; SP: subplate neuron. SP and EC are not shown in the figure due to space limit. **b,** Because of the existence of a large mass of nIPC on the map, trajectories built with only CytoTRACE **(i)** or RNA velocity **(ii)** resulted in unrealistic 'sink' and 'saddle' on the map. In contrast, *EpiTrace* age **(iii)** provided a unidirectional reference of time to reveal that the 'sink' nIPC population is mitotically active to resolve the 'nIPC stall', avoiding trajectories to be attracted to the human-specific expanded neurogenic cell population. Likewise, *EpiTrace* hybrid kernel revealed In3 as an inhibitory neuron at immature state and is capable to undergo further division, similar to reported DENs neurons. **c,** Between the branched nIPC trajectories, NR2F1 is most significantly differentially expressed and is associated with significant TFBS chromatin accessibility changes. **d,e,** NR2F1 is functionally switched into a transcriptional repressor in $NR2F1^+$ nIPC as judged by the chromatin accessibility of its TFBS **(d)** compared with its expression **(e)**. **f-h,** Hence, nIPC could be further separated into $NR2F1^+$ and $LMO3^+$ ($NR2F1^-$) **(f)** sub-populations. The two subpopulations showed different mitotic activity. $NR2F1^+$ nIPC divides faster as indicated by higher *EpiTrace* age **(g)**, higher cell number, and lost of stemness as compared to $LMO3^+$ nIPC **(h)**. It is mostly making assymetric progenitor-terminal divisions across cell age whilst $LMO3^+$ nIPC turns to produce terminal-terminal divisions as it ages, suggesting that $NR2F1^+$ nIPC clones support earlier neurogenesis and $LMO3^+/NR2F1^-$ clones expand relatively later. This result agrees well with human mutation phenotype and provides a putative explaination to the NR2F1-guided gyrification in human[54]. **i,** Mitotic clock (*EpiTrace*) and differentiation potential (CytoTRACE) of the same cell in scMultiome-seq. CytoTRACE score is reversed to show differentiation from left to right to facilitate comparison with *EpiTrace*. Notably, nIPC/GluN1 is mitotically old but maintained a

progenitor state and high differentiation potential, whilst the maturity of terminally differentated neurons is negatively correlated with their age, suggests that maturation is a function of post-mitotic latency. **j,** Excitatory neuron phylogeny built with mitotic clock, showing that GluN4/GluN5 are likely direct, early-born progenies of RG whilst GluN2/GluN3 are likely late-born, immature progenies of nIPC. *EpiTrace* model is concordant with state-of-the-art knowledge of neurogenesis and highlights post-mitotic maturation process of neuron. **k,** Overall model of corticogenesis in the light of *EpiTrace*. These results indicate a long post-mitotic maturation (differentiation) of neurons, revealed the molecular mechanism of NR2F1 controlling human nIPC assymetrical division to result in cortical gyrification, and suggested late-born nature of inhibitory neurons. Data source: Trevino *et al.*[50].

Figure 6. Single cell phylogeny on mitotic age loci indicates de-differentiation driven by master transcription factor in translocation renal cell carcinoma. a, UMAP projection of scATAC data of normal human kidney sample from biopsies. LEUK: leukocytes; ENDO: endothelial cells; MES: mesangial cells; PEC: parietal epithelioid cells; PODO: podocytes; PT: proximal tubule; PTPL: proximal tubule progenitor-like cell; LOH: loop-of-Henri; DCT: distal collection tubule; CD_PC: collection duct principal cell; CD_ICA: collection duct intercalating cell. **b,** Single cell phylogeny built with chromatin accessibility on ClockDML (*EpiTrace* phylogeny) suggests an endothelial origin of kidney tubules and delineates cell-type-specific generation cascade during nephrogenesis. **c,** Distribution of *EpiTrace* age for each cell type suggesting a distal-to-proximal genesis cascade of nephron tubules with a late expansion of proximal tubule. **d,** Cell age distribution of PT cells (colored by type) (top), and ATAC peak activity across cell age (bottom) in PT cells, showing clear segregation of peaks opened in progenitor or differentiated PT cells. Translocation renal cell carcinoma (TRCC) driver gene TFEB is specifically activated in progenitor cell and shows age-dependent decrease of activity. In contrast, differentiated cell-specific, age-dependent increased peaks are associated with hereditary renal dysgenesis (CAKUT) genes FGF8, FGFR2, SLIT3, GDNF and NHS. These results suggest that CAKUT is linked to genes functioning in terminal proximal tubule cell fate determination and function, whilst TRCC oncogenesis is linked to the mis-expression of progenitor-specific transcription factor.

Figure 7. Early-on branching evolution of glioblastoma clonal evolution tracked by single cell mitotic age. a, UMAP of scATAC data from GBM sample CGY4349. Cell clusters are labelled (Numbers). **b-e,** Per-cell-cluster EpiTrace mitotic clock age **(b)** and MDM4 **(c)**, EGFR **(d)**, PDGFRA **(e)** promoter chromatin accessibility (ChrAcc). Note that MDM4 amplification is further gained in younger clones 0/3. **f,** *EpiTrace* age of cells with normal karyotype, $MDM4^+$, $MDM4^+/EGFR^+/PDGFRA^+$, $MDM4^+/EGFR^+$, and $MDM4^+/PDGFRA^+$. **g,** *EpiTrace* phylogeny, clonal composition (pie charts on the right), showing origin pre-malignant cell cluster (7) evolved into initial malignant clonal expansion with MDM4 amplification, followed by a malignant $MDM4^+$ population which is double-positive with PDGFRA and EGFR

amplifications. In later stage, tumor clones split the oncogenic amplification and grow into either PDGFRA^{+} (5) or EGFR^{+} (6/0/3) populations. **h,** Schematic evolution trajectory.

Extended Data Figure Legends

Extended Data Figure 1. Using DNA methylation level on the ClockDML set used in this study to predict donor age from PBMC bisulfite-capture sequencing. a-d, The DNAm level were extracted for ClockDML loci and fit to predict donor age (x axis) in the training cohort with TIMEseq method **(a)** or GLMnet method **(c)**. Validation was done with an additional cohort **(b, d)**. Correlation coefficients: TIMEseq: 0.9751 (Training) / 0.7998 (Validation); GLMnet: 0.9999 (Training) / 0.8553 (Validation). Sequencing datapoints were colored according to their mean coverage on ClockDML.

Extended Data Figure 2. ChrAcc pattern on ClockDML gradually converges during cell development. a-d, Bulk ATAC-seq of Fluorescence-Activated-Cell-Sorting(FACS)-sorted blood cells overlapping ClockDML region **(a)** were used to compute the inter-sample-variation of ClockDML ChrAcc **(b)**, per-sample normalized ClockDML ChrAcc **(c)**, and per-sample ClockDML ChrAcc entropy **(d)**. **e-f,** Chromatin accessibility pattern is more chaotic in progenitor cells and gradually converged to fixed pattern in terminal cells, as evident in the per-sample cross-peak coefficient of variation **(e)** and the increase of percentage of reads originating from ClockDML-associated peaks **(f)**.

Extended Data Figure 3. Undersampling in single cell sequencing results in reverse of observable total ChrAcc on ClockDML peaks compared to bulk ATAC sequencing. a, Schematic of chromatin accessibility on ClockDML in early, intermediate or late-born cells. Though accessible loci are more in early cells, the completely opened loci are enriched in terminal cells. **b,** Pseudobulk scATAC signals on a ClockDML around RB1CC1 from human placental trophoblast stem cell (TSC), villous cytotrophoblast (VCT) and extravillous trophoblast (EVT). Completely opened loci are schematically shown above the gray threshold line. Data from Gong et al, unpublished.

Extended Data Figure 4. ClockDML ChrAcc in bulk samples. a, Smoothened ChrAcc on FACS-sorted HSC. **b,** Smoothened ChrAcc on lymphoblastoid, iPSC, and iPSC-Cardiomyocyte.

Extended Data Figure 5. Bulk ATAC cell age estimation using reverse EpiTrace Age. a, Schematic diagram of hematopoiesis. **b,** Sample age (x-axis, reverse *EpiTrace* age) of bulk sample ATAC-seq data from FACS-sorted normal and leukemic cells. **c,** *EpiTrace* phylogeny of hematopoietic sub-lineages. **d,** Reverse *EpiTrace* age of bulk ATAC-seq of lymphoblastoid cell (LCL), derived iPSC (iPSC), and iPSC-derived cardiomyocytes (iPSC-CM). Samples from the same donor across different time points are grouped (lines). **e,** Reverse *EpiTrace* age of FACS-sorted, naïve- or non-naïve CD8 cells from PBMC or tumor tissue.

Extended Data Figure 6. Details of 8-cell-like induction scMultiome ATAC dataset. a, scMultiome ATAC UMAP projection, cell clusters are labelled. Trajectories between clusters are drawn. 8-cell-like (8CL) cluster are colored in blue. **b,** *EpiTrace* age prediction of single cells in each cluster, ordered by mean *EpiTrace* age. **c,** DPPA3 enhancer activation in each cluster. Clusters are labelled as 'Primed' (ground-zero), 'early', 'intermediate' and 'terminal', according to their position.

Extended Data Figure 7. UMAP projection of scMultiome-sequenced D17 culture with single cell evolution trajectories built with RNA velocity. Cells are colored with stemness (CytoTRACE) score and annotated cell types are labelled on the projection. The gradient of CytoTRACE score does not explicitly points to the target 8-cell-like state (8CLC). RNA velocity-built evolution trajectories end with expected terminal states, but start from a differentiated cell cluster, which is inconsistent with biological reality.

Extended Data Figure 8. Erythroid lineage shows error of EpiTrace cell age estimation as a result of genome-wide chromatin accessibility closing. a, Cell collection day projected to UMAP. **b,** *EpiTrace* age of single cell on UMAP. **c,** Phenotypic pseudotime of erythroid lineage. **d,** Phenotypic pseudotime of myeloid lineage. **e-h,** Examples of cell clones initially differentiating into myeloid compartment at uninduced state (**e-f)** but forced to be differentiated into erythroid compartment at later timepoint **(g-h)**. Erythroid lineage differentiation is accompanied with a significant reduction of *EpiTrace* age, probably due to genome-wide chromatin accessibility closing prior to terminal erythrogenesis.

Extended Data Figure 9. Age shift in T cell clones following induced differentiation.

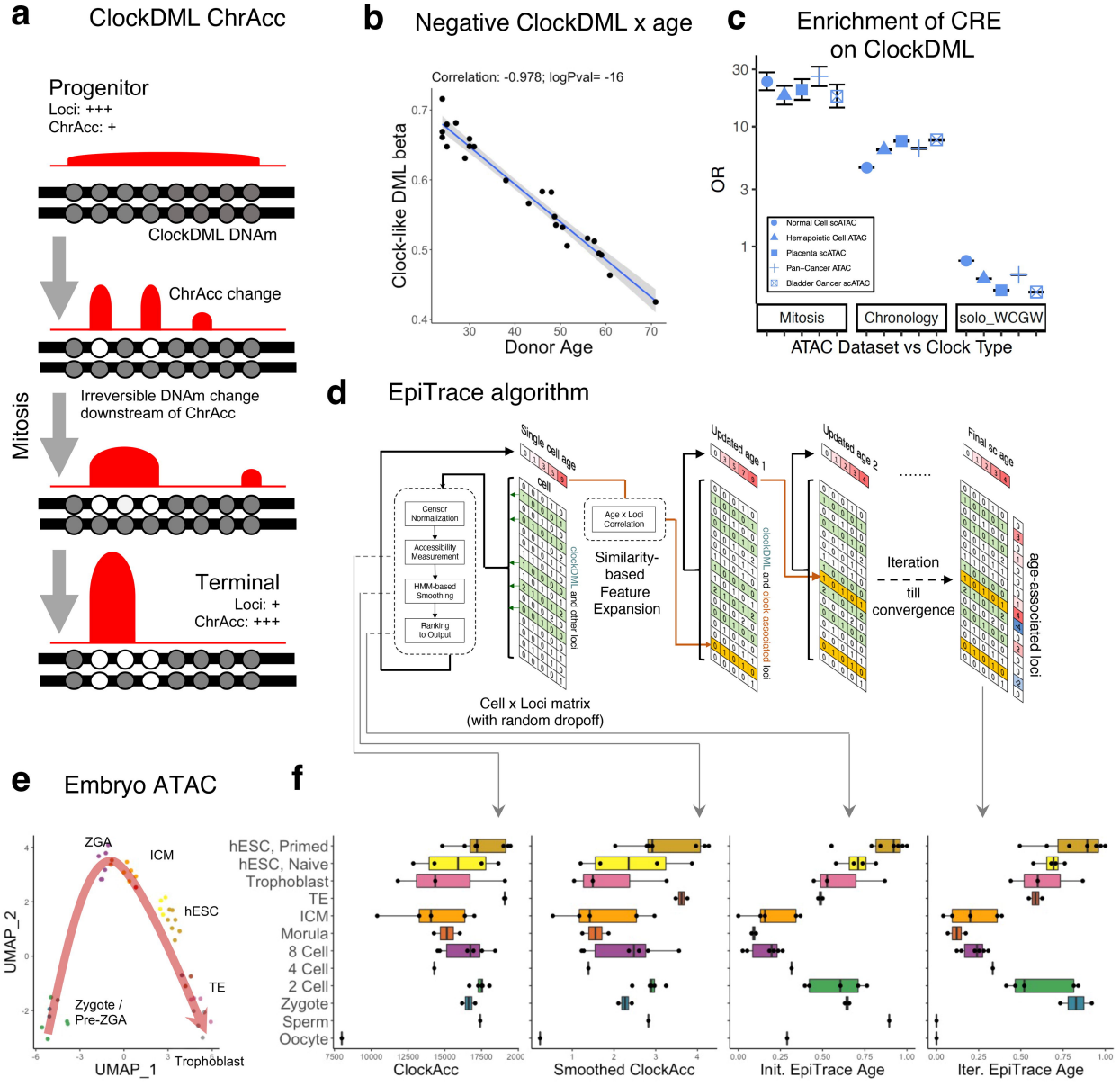
Extended Data Figure 10. Cell lineage inferred by phenotype-only method (top) or considering both phenotype and mitosis age. A standard division time for each mitosis is assumed. Assuming the current time is n mitosis (n), and the lineage begins to divide from a single radial glia (RG, left-most), which undergone k+1 symmetric divisions until the first assymetric division (k+2) to give rise to a RG and a subplate neuron (SP), then the birth-time of SP neuron is (k+2), and the post-mitotic maturation time for SP neuron is n-(k+2). The logic follows for other glutaminergic neurons shown in the schematic figure.

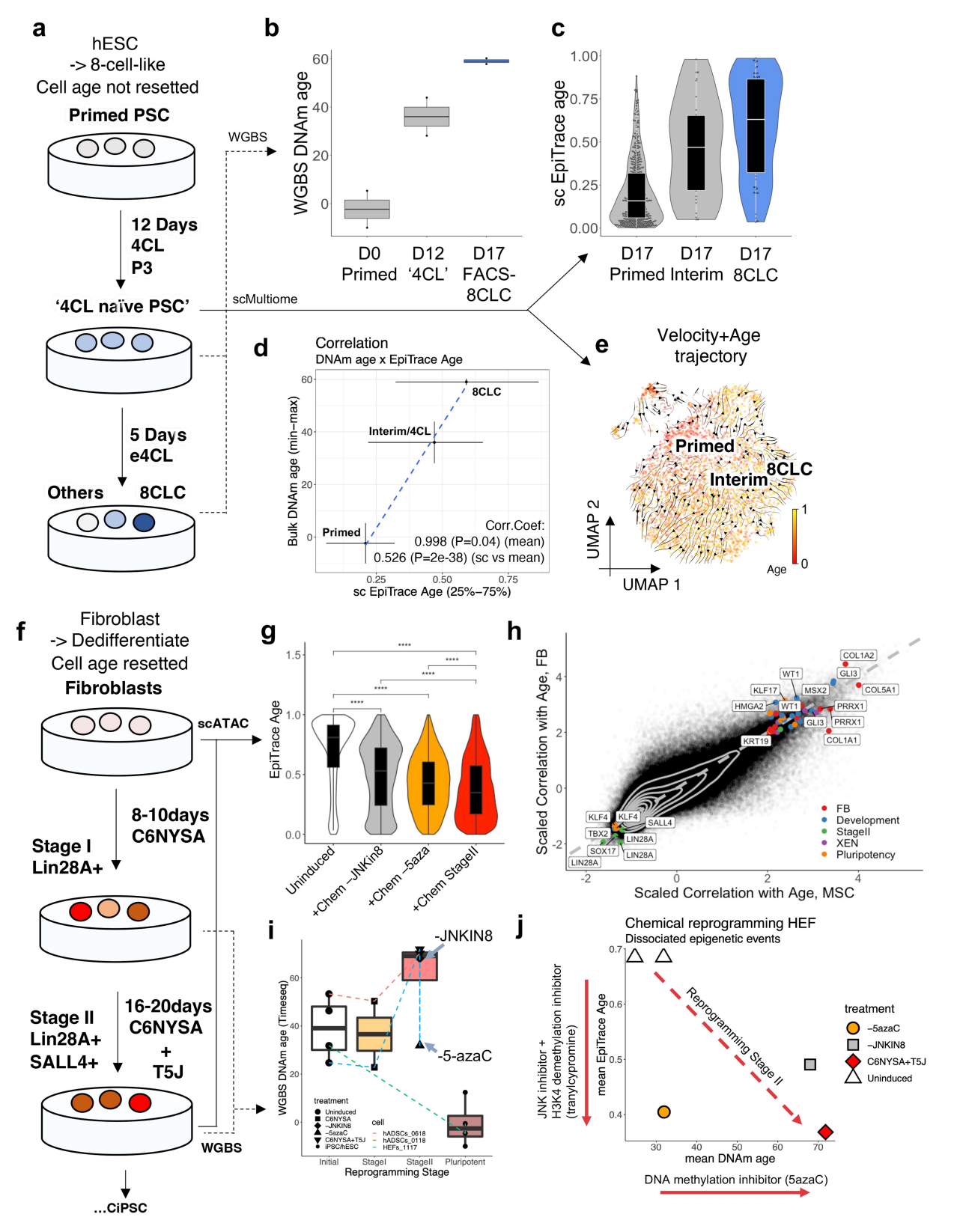
Extended Data Figure 11. Additional schematic drawing to show correlation between anatomical positioning and EpiTrace phylogeny of kidney cells.

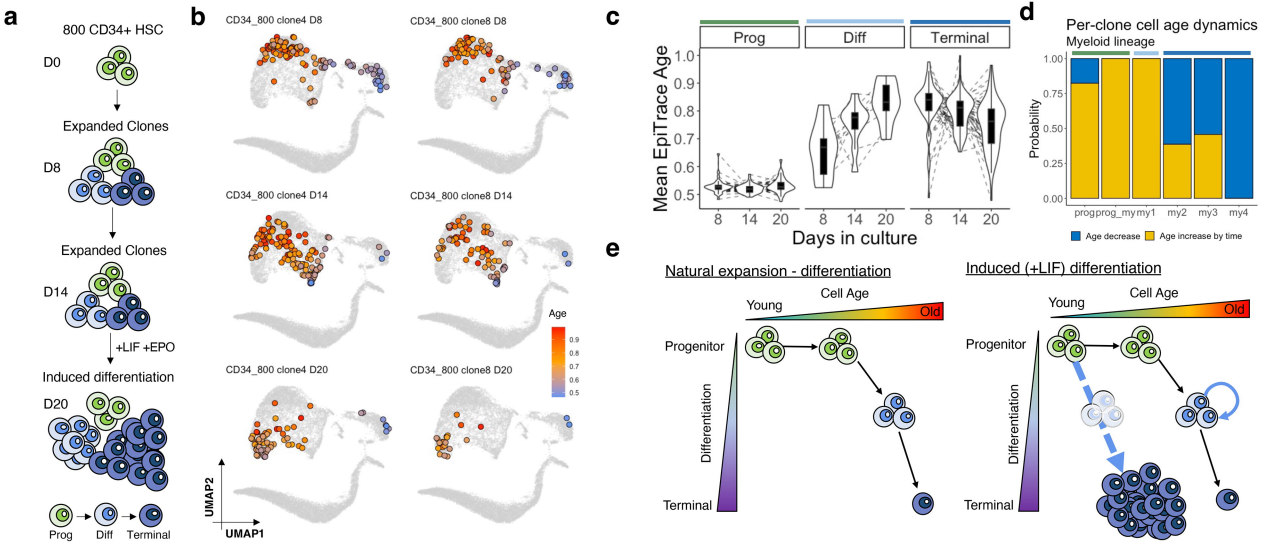
Extended Data Figure 12. Branching evolution of EGFR$^+$ and PDGFRA$^+$ clusters from a double-positive EGFR$^+$/PDGFRA$^+$ cancer clone. Scatter plot (left) shows EGFR and PDGFRA peak height of each single cell within the designated cluster. Single cell EGFR peak height (y axis) as a function of *EpiTrace* age (x axis) was

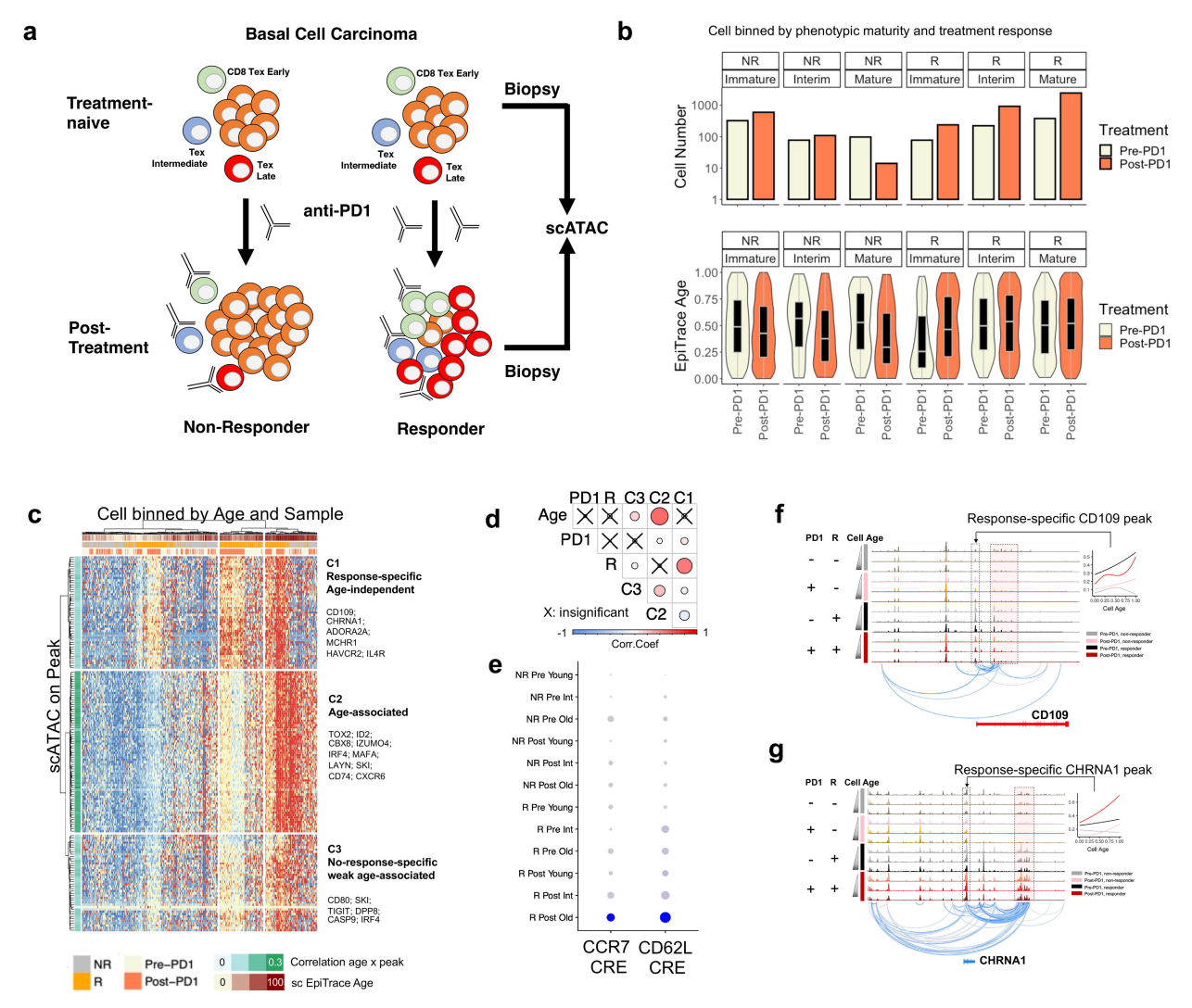
shown on the right, superimposed with regressed EGFR peak height over time for $EGFR^+$ (red) and $PDGFRA^+$ (black) clones. Branching evolution starts around *EpiTrace* age 0.3 and results in either $EGFR^+$ or $PDGFRA^+$ clones, as most evident in cluster 6 (consisting of both $EGFR^+$ and $PDGFRA^+$ cells).

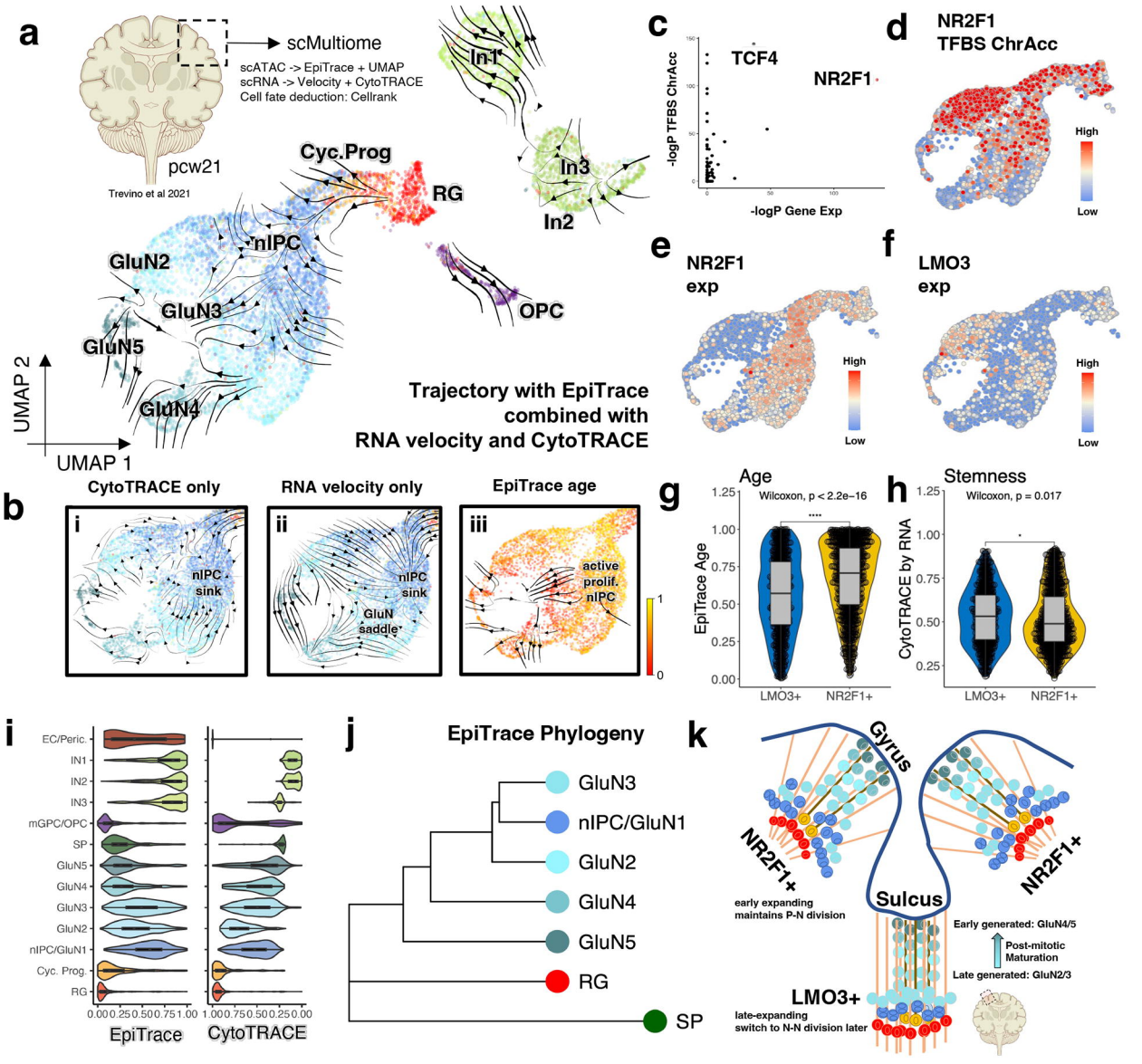
Extended Data Table 1. ClockDML reference set used in the study (in hg19 coordinates).

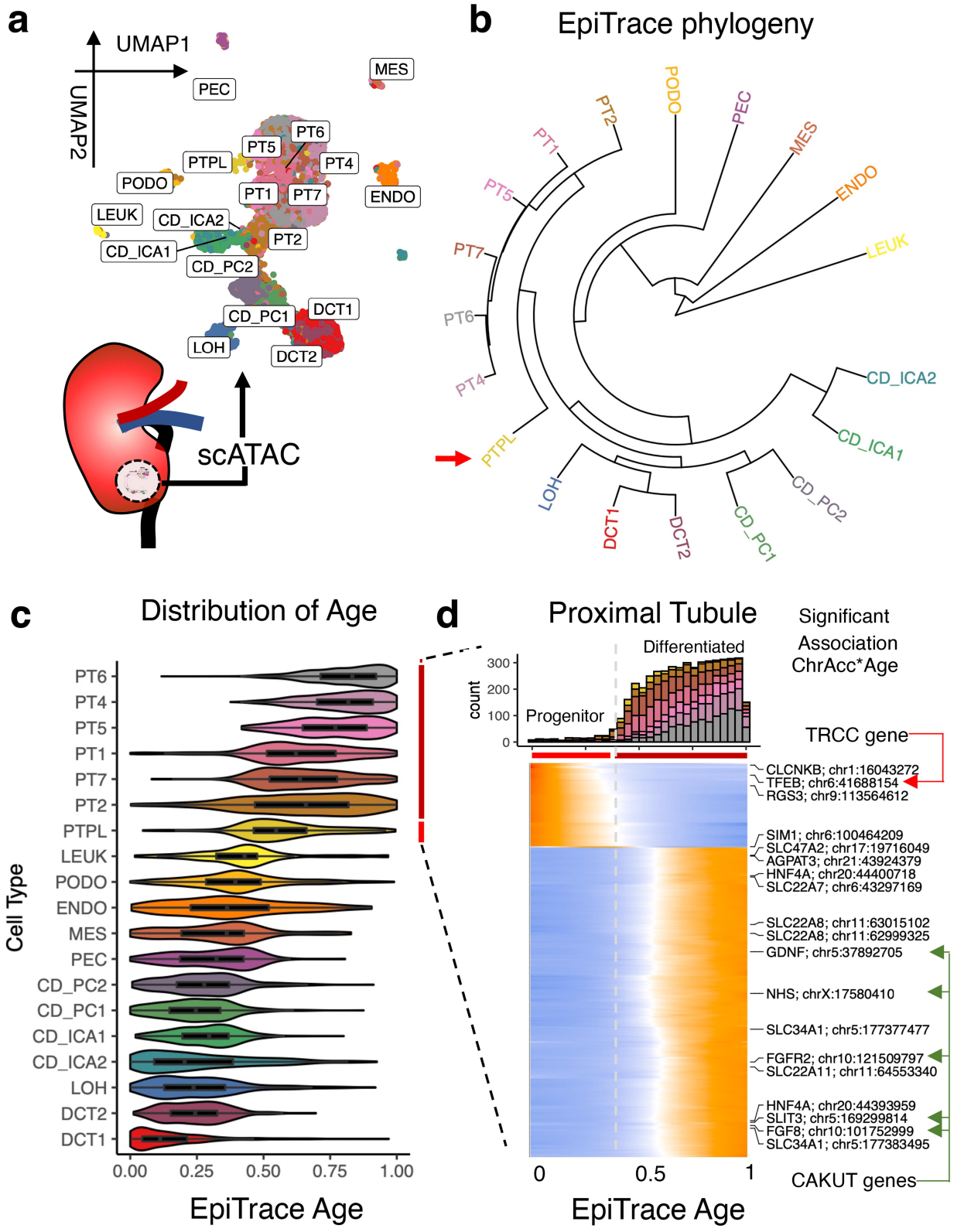

a
ClockDML ChrAcc
Progenitor
Loci: +++
ChrAcc: +
ClockDML DNAm
ChrAcc change
Mitosis
Irreversible DNAm change
downstream of ChrAcc
Terminal
Loci: +
ChrAcc: +++
b
Negative ClockDML x age
Correlation: -0.978; logPval= -16
Clock-like DML beta
Donor Age
c
Enrichment of CRE
on ClockDML
OR
Normal Cell scATAC
Hemapoietic Cell ATAC
Placenta scATAC
Pan-Cancer ATAC
Bladder Cancer scATAC
Mitosis
Chronology
solo_WCGW
ATAC Dataset vs Clock Type
d
EpiTrace algorithm
Single cell age
Updated age 1
Updated age 2
Final sc age
Censor Normalization
Accessibility Measurement
HMM-based Smoothing
Ranking to Output
Age x Loci Correlation
Similarity-based Feature Expansion
clockDML and other loci
clockDML and clock-associated loci
Iteration till convergence
age-associated loci
Cell x Loci matrix
(with random dropoff)
e
Embryo ATAC
ZGA
ICM
hESC
TE
Zygote /
Pre-ZGA
Trophoblast
UMAP_1
UMAP_2
f
hESC, Primed
hESC, Naive
Trophoblast
TE
ICM
Morula
8 Cell
4 Cell
2 Cell
Zygote
Sperm
Oocyte
ClockAcc
Smoothed ClockAcc
Init. EpiTrace Age
Iter. EpiTrace Age

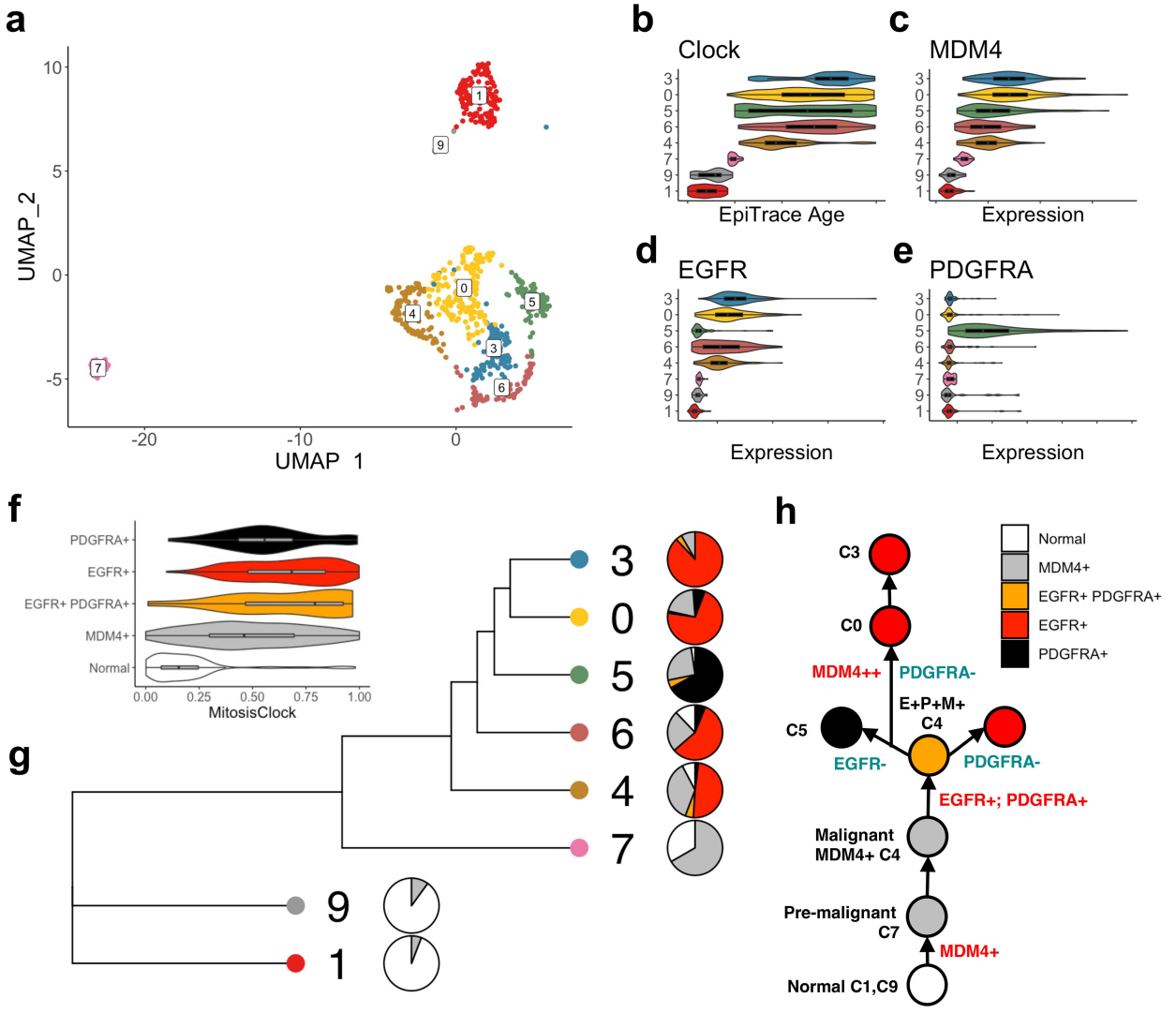

a
hESC
-> 8-cell-like
Cell age not resetted
Primed PSC
12 Days
4CL
P3
'4CL naïve PSC'
5 Days
e4CL
Others
8CLC
WGBS
scMultiome
b
WGBS DNAm age
D0 Primed
D12 '4CL'
D17 FACS-8CLC
c
sc EpiTrace age
D17 Primed
D17 Interim
D17 8CLC
d
Correlation
DNAm age x EpiTrace Age
Bulk DNAm age (min–max)
sc EpiTrace Age (25%–75%)
8CLC
Interim/4CL
Primed
Corr.Coef:
0.998 (P=0.04) (mean)
0.526 (P=2e-38) (sc vs mean)
e
Velocity+Age
trajectory
Primed
Interim
8CLC
UMAP 2
UMAP 1
Age
f
Fibroblast
-> Dedifferentiate
Cell age resetted
Fibroblasts
scATAC
8-10days
C6NYSA
Stage I
Lin28A+
Stage II
Lin28A+
SALL4+
16-20days
C6NYSA
+
T5J
WGBS
...CiPSC
g
EpiTrace Age
Uninduced
+Chem –JNKin8
+Chem –5aza
+Chem StageII
h
Scaled Correlation with Age, FB
Scaled Correlation with Age, MSC
COL1A2
GLI3
COL5A1
WT1
KLF17
MSX2
HMGA2
PRRX1
KRT19
COL1A1
KLF4
SALL4
TBX2
LIN28A
SOX17
FB
Development
StageII
XEN
Pluripotency
i
WGBS DNAm age (Timeseq)
-JNKIN8
-5-azaC
treatment
Uninduced
C6NYSA
–JNKIN8
–5azaC
C6NYSA+T5J
iPSC/hESC
cell
hADSCs_0618
hADSCs_0118
HEFs_1117
Initial
StageI
StageII
Pluripotent
Reprogramming Stage
j
Chemical reprogramming HEF
Dissociated epigenetic events
mean EpiTrace Age
mean DNAm age
Reprogramming Stage II
JNK inhibitor +
H3K4 demethylation inhibitor
(tranylcypromine)
DNA methylation inhibitor (5azaC)
treatment
–5azaC
–JNKIN8
C6NYSA+T5J
Uninduced


a
800 CD34+ HSC
D0
Expanded Clones
D8
Expanded Clones
D14
+LIF +EPO
Induced differentiation
D20
Prog
Diff
Terminal
b
CD34_800 clone4 D8
CD34_800 clone8 D8
CD34_800 clone4 D14
CD34_800 clone8 D14
CD34_800 clone4 D20
CD34_800 clone8 D20
Age
0.9
0.8
0.7
0.6
0.5
UMAP2
UMAP1
c
Prog
Diff
Terminal
Mean EpiTrace Age
1.0
0.9
0.8
0.7
0.6
0.5
8
14
20
Days in culture
d
Per-clone cell age dynamics
Myeloid lineage
Probability
1.00
0.75
0.50
0.25
0.00
prog
prog_my
my1
my2
my3
my4
Age decrease
Age increase by time
e
Natural expansion - differentiation
Cell Age
Young
Old
Progenitor
Differentiation
Terminal
Induced (+LIF) differentiation
Cell Age
Young
Old
Progenitor
Differentiation
Terminal


a
Basal Cell Carcinoma
Treatment-naive
CD8 Tex Early
Tex Intermediate
Tex Late
Biopsy
anti-PD1
scATAC
Post-Treatment
Biopsy
Non-Responder
Responder
b
Cell binned by phenotypic maturity and treatment response
NR Immature
NR Interim
NR Mature
R Immature
R Interim
R Mature
Cell Number
EpiTrace Age
Treatment
Pre-PD1
Post-PD1
c
Cell binned by Age and Sample
scATAC on Peak
C1
Response-specific
Age-independent
CD109;
CHRNA1;
ADORA2A;
MCHR1
HAVCR2; IL4R
C2
Age-associated
TOX2; ID2;
CBX8; IZUMO4;
IRF4; MAFA;
LAYN; SKI;
CD74; CXCR6
C3
No-response-specific
weak age-associated
CD80; SKI;
TIGIT; DPP8;
CASP9; IRF4
NR
R
Pre–PD1
Post–PD1
Correlation age x peak
sc EpiTrace Age
d
PD1 R C3 C2 C1
Age
PD1
R
C3
C2
X: insignificant
Corr.Coef
e
NR Pre Young
NR Pre Int
NR Pre Old
NR Post Young
NR Post Int
NR Post Old
R Pre Young
R Pre Int
R Pre Old
R Post Young
R Post Int
R Post Old
CCR7 CRE
CD62L CRE
f
Response-specific CD109 peak
PD1 R Cell Age
Cell Age
CD109
g
Response-specific CHRNA1 peak
PD1 R Cell Age
Cell Age
CHRNA1


a
scMultiome
scATAC -> EpiTrace + UMAP
scRNA -> Velocity + CytoTRACE
Cell fate deduction: Cellrank
pcw21
Trevino et al 2021
Cyc.Prog
RG
In1
In3
In2
nIPC
GluN2
GluN3
GluN5
GluN4
OPC
UMAP 2
UMAP 1
Trajectory with EpiTrace
combined with
RNA velocity and CytoTRACE
b
CytoTRACE only
RNA velocity only
EpiTrace age
i
ii
iii
nIPC
sink
GluN
saddle
active
prolif.
nIPC
1
0
c
TCF4
NR2F1
-logP TFBS ChrAcc
-logP Gene Exp
d
NR2F1
TFBS ChrAcc
High
Low
e
NR2F1
exp
High
Low
f
LMO3
exp
High
Low
g
Age
Wilcoxon, p < 2.2e−16
EpiTrace Age
LMO3+
NR2F1+
h
Stemness
Wilcoxon, p = 0.017
CytoTRACE by RNA
LMO3+
NR2F1+
i
EC/Peric.
IN1
IN2
IN3
mGPC/OPC
SP
GluN5
GluN4
GluN3
GluN2
nIPC/GluN1
Cyc. Prog.
RG
EpiTrace
CytoTRACE
j
EpiTrace Phylogeny
GluN3
nIPC/GluN1
GluN2
GluN4
GluN5
RG
SP
k
Gyrus
Sulcus
NR2F1+
NR2F1+
early expanding
maintains P-N division
Early generated: GluN4/5
Post-mitotic
Maturation
Late generated: GluN2/3
LMO3+
late-expanding
switch to N-N division later


a
UMAP1
UMAP2
PEC
MES
PT6
PT5
PTPL
PT4
PODO
PT1
PT7
ENDO
LEUK
CD_ICA2
PT2
CD_ICA1
CD_PC2
CD_PC1
DCT1
LOH
DCT2
scATAC
b
EpiTrace phylogeny
PODO
PT2
PEC
PT1
MES
PT5
ENDO
PT7
LEUK
PT6
PT4
CD_ICA2
PTPL
CD_ICA1
LOH
CD_PC2
DCT1
DCT2
CD_PC1
c
Distribution of Age
Cell Type
PT6
PT4
PT5
PT1
PT7
PT2
PTPL
LEUK
PODO
ENDO
MES
PEC
CD_PC2
CD_PC1
CD_ICA1
CD_ICA2
LOH
DCT2
DCT1
0.00
0.25
0.50
0.75
1.00
EpiTrace Age
d
Proximal Tubule
Significant Association ChrAcc*Age
Differentiated
Progenitor
count
300
200
100
0
TRCC gene
CLCNKB; chr1:16043272
TFEB; chr6:41688154
RGS3; chr9:113564612
SIM1; chr6:100464209
SLC47A2; chr17:19716049
AGPAT3; chr21:43924379
HNF4A; chr20:44400718
SLC22A7; chr6:43297169
SLC22A8; chr11:63015102
SLC22A8; chr11:62999325
GDNF; chr5:37892705
NHS; chrX:17580410
SLC34A1; chr5:177377477
FGFR2; chr10:121509797
SLC22A11; chr11:64553340
HNF4A; chr20:44393959
SLIT3; chr5:169299814
FGF8; chr10:101752999
SLC34A1; chr5:177383495
CAKUT genes
0
0.5
1
EpiTrace Age


a
UMAP_2
UMAP_1
b
Clock
EpiTrace Age
c
MDM4
Expression
d
EGFR
Expression
e
PDGFRA
Expression
f
PDGFRA+
EGFR+
EGFR+ PDGFRA+
MDM4+
Normal
MitosisClock
g
h
Normal
MDM4+
EGFR+ PDGFRA+
EGFR+
PDGFRA+
C3
C0
MDM4++
PDGFRA-
E+P+M+
C4
C5
EGFR-
PDGFRA-
EGFR+; PDGFRA+
Malignant
MDM4+ C4
Pre-malignant
C7
MDM4+
Normal C1,C9