Cell Tree Rings: the shape of somatic evolution as a human aging timer

Attila Csordas[1,2], Botond Sipos[1], Terezia Kurucova[3,4], Andrea Volfova[5], Frantisek Zamola[5], Boris Tichy[3], Damien G Hicks[1,6]

[1] AgeCurve Limited, Cambridge, CB2 1SD, UK
[2] Doctoral School of Clinical Medicine, University of Szeged, H–6720, HU
[3] CEITEC – Central European Institute of Technology, Masaryk University, 62500 Brno, CZ
[4]Department of Experimental Biology, Faculty of Science, Masaryk University, 62500, Brno, CZ
[5] HealthyLongevity.clinic Inc, 540 University Ave, Palo Alto, CA 94301, US
[6] Swinburne University of Technology, Hawthorn, VIC 3122, AU

Abstract

Biological age is typically estimated using biomarkers whose states have been observed to correlate with chronological age. A persistent limitation of such aging clocks is that it is difficult to establish how the biomarker states are related to the mechanisms of aging. Somatic mutations could potentially form the basis for a more fundamental aging clock since the mutations are both markers and drivers of aging and have a natural timescale. Cell lineage trees inferred from these mutations reflect the somatic evolutionary process and thus, it has been conjectured, the aging status of the body. Such a timer has been impractical thus far, however, because detection of somatic variants in single cells presents a significant technological challenge.
Here we show that somatic mutations detected using single-cell RNA sequencing (scRNAseq) from hundreds of cells can be used to construct a cell lineage tree whose shape correlates with chronological age. De novo single-nucleotide variants (SNVs) are detected in human peripheral blood mononuclear cells using a modified protocol. Penalized multiple regression is used to select from over 30 possible metrics characterizing the shape of the phylogenetic tree resulting in a Pearson correlation of 0.8 between predicted and chronological age and a median absolute error less than 6 years. The geometry of the cell lineage tree records the structure of somatic evolution in the individual and represents a new modality of aging timer. In addition to providing a single number for biological age, it unveils a temporal history of the aging process, revealing how clonal structure evolves over life span. Cell Tree Rings complements existing aging clocks and may help reduce the current uncertainty in the assessment of geroprotective trials.

Introduction

Aging refers to the systematic decline in cellular and organismal function over time. The ubiquity of age-related disease makes chronological age the single most important risk factor for morbidity and mortality [Partridge et al, 2018]. Interventions to

slow, delay or even reverse the aging process thus have the potential to mitigate multiple age-related pathologies [Kaeberlein, 2017].

To quantify the effectiveness of such interventions it is necessary to have a reliable measure of biological age. Aging timers, or clocks, accomplish this by using specific biomarkers whose states change systematically with chronological age. A variety of biomarker modalities have been studied, particularly epigenetic, but also transcriptomic, proteomic and metabolomic, among others [Rutledge et al., 2022; Macdonald-Dunlop et al., 2022]. A necessary step in the development of current aging clocks is to show that the chosen biomarker states are associated with chronological age across a population [Horvath, Raj 2018]. This correlation captures the average changes over lifespan and establishes a baseline to which individuals can be compared. A desirable property of these biomarker timers is that they be directly linked to the hallmarks of aging [López-Otín et al, 2013]. This potentially allows the biomarker states to be interpreted in terms of the mechanisms of aging.

Genome instability due to somatic mutations is the first hallmark of aging [López-Otín et al, 2013]. In blood, mutations can lead to somatic mosaicism and eventually clonal hematopoiesis, where cell populations harbouring particular allele variants outgrow others. Animal models of clonal hematopoiesis have been shown to contribute to disease progression [Evans and Walsh, 2022]. More generally, diseases characterized by accelerated aging typically involve the increased accumulation of DNA damage [Lodato et al., 2018] Given its importance as a driver of aging, it would seem that somatic evolution could form the basis for a new type of aging timer.

Somatic mutations (single-nucleotide variants, SNVs, and copy-number variants, CNVs) are naturally-occurring barcodes [Sankaran et al, 2022] that enable phylogenetic inference of cell lineage trees (cell trees from now on). Cell trees are a representation of the mitotic branching order and clonal structure of a sampled cell population [Salipante and Horwitz, 2006, Wasserstrom et al, 2008]. These partial cell trees are subtrees of the whole organismal cell lineage tree, which in an adult human consists of tens of trillions of cells [Sender et al, 2016]. The *shape* of a tree refers to the ordering and length of its branches and reflects the clonal structure and evolutionary distances between cells.

The central conjecture behind our proposed aging timer is that the shape of cell trees is a representation of the biological aging process [Csordas, 2019]. There are two reasons for this hypothesis. The first is that phylogenetic systematics has long shown

how genetic distances between species existing today reflect evolutionary changes in the past. It is reasonable then to expect that genetic distances between single cells can be used to infer the somatic evolutionary history of cells, a driver and indicator of aging. The second is that biomedical life history can leave its imprint on the cell tree [Stadler et al, 2021], providing a record of major transitions in the aging process. An additional benefit of cell trees is that they provide an intuitively appealing representation of the dynamics of aging that naturally lends itself to interpretation.

Using human peripheral blood cells from healthy individuals (n=18, age range 21-82 years of age) we have developed a new aging timer called Cell Tree Rings (CTR) with the following characteristics:

1. Naturally occurring somatic single nucleotide variants (SNVs) are used to build cell trees using standard phylogenetic algorithms,
2. SNVs are called directly and de novo from scRNA-seq data from hundreds or thousands of cells,
3. A comprehensive set of tree metrics is used to identify aspects of tree shape that are associated with chronological age using a penalized multiple regression model.

Results

Cell trees

Approximately 1400 cells are recorded from each of the 18 individuals in the study. 25 pseudo-replicate trees were generated for each individual. Each tree was constructed from a random sample of 700 cells from that individual. Phylogenetic trees were inferred using the distance matrix algorithm UPGMA. Figure 1 shows one pseudo-replicate tree from each individual. The circular rendering, which places the root at the centre and the cells around the perimeter, provided inspiration for the name `Cell Tree Rings'.

Figure 1: A single cell tree from each of the 18 participants in the study. To estimate uncertainty in the tree shape metrics, a total of 25 pseudo-replicate trees were inferred for each participant. Each tree is constructed from a random sample of 700 cells out of the ~1400 from that individual.

Cell tree metrics

The central hypothesis of the study is that the shape of cell trees is a measure of biological age. Here shape refers to the combination of topology (branching order) and branch lengths. Topology, in the case of cell lineage trees, corresponds to branching patterns of mitotic division in somatic cells. Branch lengths represent the amount of evolutionary change and are usually defined as the product of mutation rates and a suitable unit of time. Various tree statistics capture either topology only or branch length only information, or they can capture a combination of both.

We have applied a set of 30+ tree metrics to characterize cell trees built from somatic mutations from human peripheral blood mononuclear cells. This set of measures comprises both traditional tree metrics used in phylogenetics and some that were developed specifically for this study.

Identifying tree features with high biological signal/technical variation noise ratio

A necessary condition for a reliable aging timer is that the biological variation between individuals over a wide range of ages should be considerably larger than the technical variation within individuals. Metrics for which this is not the case are not useful as an aging timer and are removed.

For every individual 25 pseudo-replicate trees have been produced using random subsets, with replacement, of 700 cellular barcodes.

The biological to technical variation is estimated notionally using an F-ratio. Note that, because the pseudo-replicates are only partially independent, the value of the F-ratio should not be interpreted formally as such; it is the comparison across metrics that is of interest. A rough cut-off for eliminating 'noisy' metrics is shown on Figure 2, indicating which of the 25 tree metrics passed this test. Note that metrics which are included are not necessarily good timers since they need to also contribute to the correlation with age.

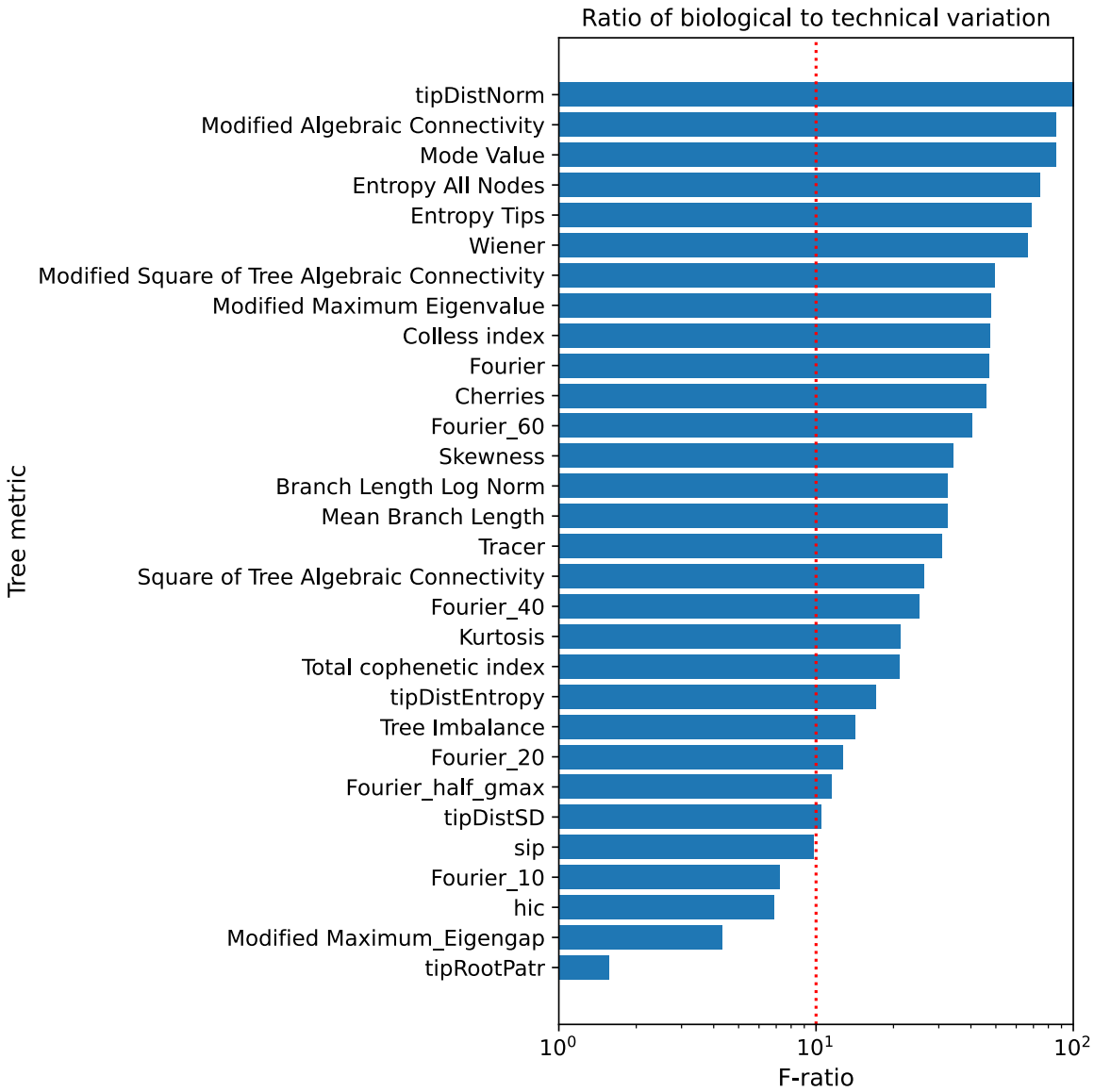

Figure 2: F-ratios of tree metrics using UPGMA cell trees from 25 pseudo-replicate trees with 700 cells at their tips. Metrics below the cut-off ratio of 10 are eliminated.

Model building and Prediction Results

For model selection, penalized regression using the Lasso, cross-validated using Least Angle Regression, has been applied. Since Lasso is unstable with highly correlated predictor variables [Hastie et al, 2015, p56], the 25 tree features selected in the previous high biological signal/technical variation noise ratio filtering step were subsequently checked for multicollinearity by computing their variance inflation factors (VIF). A rule of thumb commonly used is that predictor variables with over 5 VIF value

are excluded. This step reduced the number of tree features used for model selection to the following metrics: Mean Branch Length, Tree Imbalance, Fourier_40, Tracer and Wiener. These five features plus Sex as a binary valuable was used in the LARS-Lasso procedure, where interaction terms were allowed.

During the leave-one-out Cell Tree Age prediction step mean values for each individual are used as the predictor and response. This training data is then used to predict the Cell Tree biological age of all the replicates of the test individual left out of the training data. The final Cell Tree Age of the test individuals is established by averaging the predicted Cell Tree Ages across the replicates belonging to the same test individual. The error terms, correlation coefficient and explained variance are estimated in the test set by comparing the Cell Tree Ages to the actual chronological ages.

The best performing prediction model had only three predictor variables selected: Mean Branch Length, Tree Imbalance and Sex. Figure 3 shows a scatter plot of the best performing Cell Tree Age test prediction model on human blood samples from 18 individuals.

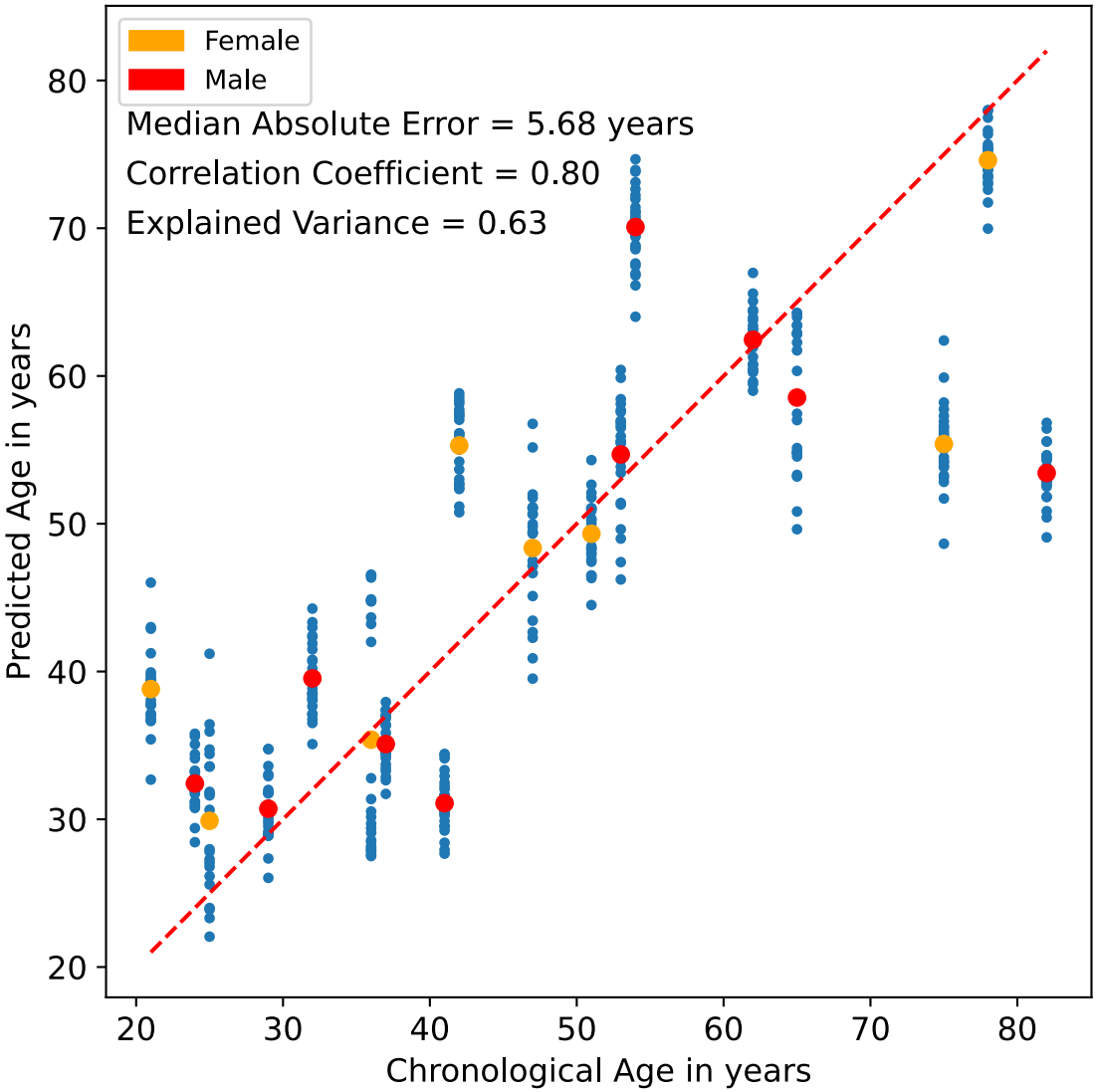

Figure 3: Scatter plot of the Cell Tree Age test prediction on 18 human blood samples from the best performing prediction model using Mean Branch Length, Tree Imbalance and Sex as predictor variables. The blue dots correspond to the predicted ages of the particular pseudo-replicate trees per test individual, and the orange and red dots represent the predicted mean ages of the 25 pseudo-replicate trees for the test individuals, females and males correspondingly. The x axis is Chronological Age in years, and the y axis shows the numerical values of the predicted Cell Tree Ages in years. The red diagonal, x=y dashed line is for display purposes only. Mean Absolute Error in years, Pearson's Correlation Coefficient and Explained Variances are shown.

A permutation test was performed by randomizing the chronological ages of the samples and comparing the regression results to the unrandomized case. Of the 100 randomized cases, none had a MAE less than that of the unrandomized case, indicating a p-value of less than 0.01.

A simpler model which includes only Mean Branch Length and Sex, without using Tree Imbalance, was still predictive, although with poorer performance. This minimal model, as we will call it, had a Median Absolute Error of 7.95 years, a Correlation Coefficient of 0.68 and an Explained Variance of 0.46.

Discussion

We have shown that cell trees constructed using SNVs from human peripheral blood mononuclear cells can reliably and reproducibly predict chronological age. The SNVs underlying these trees can be directly called from the most accessible single-cell sequencing approach, scRNA-seq.

The resulting new molecular aging timer, Cell Tree Rings, involves penalized regression of chronological age on dozens of potential cell tree metrics. The regression results show that, to predict chronological age with a median absolute error of less than 6 years, a correlation coefficient of 0.8 and with 63% of explained variance, requires only three features: Sex, the mean branch length between all the nodes in the tree, and a generalized tree imbalance factor. The minimal model, using only Sex and mean branch length, predicts chronological age with a median absolute error of ~8 years and correlation coefficient of 0.68. Such a simple model lends itself to a suggestive interpretation: the mean branch length component is connected to the linear lifelong accumulation of somatic mutations in the hematopoietic system; this has been shown spectacularly with single cell DNA sequencing of in vitro cultured hematopoietic stem and progenitor cells from the human bone marrow and peripheral blood [Mitchell et al, 2022]. This is consistent with our results using scRNAseq data from in vivo PBMCs taken fresh from blood and not cultured in vitro beforehand. Another difference between Cell Tree Rings and the in vitro scDNA-seq approach is that the PBMC's isolated in vivo contain both undifferentiated and differentiated mature white blood cells, i.e. hematopoietic stem and progenitor cells, T cells and B cells, and macrophages, among others.

Adding a generalized Tree Imbalance component makes for a better performing aging timer compared to the minimal model, with ~2 years smaller median absolute error. The Tree Imbalance variable can capture changes in the global asymmetry of the tree

shape with age. This composite variable in our current model is the product of the Kurtosis and Skewness values of the spectral density profile of the eigenvalues of the Modified Graph Laplacian of the cell trees. Both have an algebraic interpretation on simulated trees: Kurtosis can reflect even or uneven distribution of the eigenvalues and hence imbalances of the branching order of the tree with branch lengths factored in as weights, while Skewness values are indicative of the stem-to-tip structure of the tree reflecting asymmetry of the density profile. [Lewitus and Morlon, 2016]. Overall, the Tree Imbalance of the tree shape might represent a conceptually new and experimentally verifiable class of quantitative predictors of age at the level of cell trees.
We believe Cell Tree Rings to be the only existing natural barcoding approach that can build larger trees from thousands of cells using somatic mutations called de novo from scRNA-seq data and from all autosomes, sex chromosomes and the mitochondrial genome combined. This way representative cell trees can capture multiple clonal events in different parts of the accessible cellular genomes, reaching a higher resolution of the cell population history than based on targeted approaches alone.
The other advantage of Cell Tree Rings is the ability to extract both lineage histories and gene expression levels from the very same cells using only one method and lab protocol. Combining lineage and phenotypic expression information from the same sample is a considerable challenge and existing approaches offer complex solutions combining different protocols [Lähnemann et al, 2020, Sankaran et al, 2022].

Potential directions for improvement in CTR

The version of CTR reported here represents a basic proof-of-principle. Here we discuss some of the improvements envisioned. The current version of Cell Tree Rings has a considerable amount of unexplained variation with R^2 being 0.63. These results have been achieved with direct and de novo scRNA-seq alone without the aid of any bulk sequencing approach. At a technical level, bulk exome sequencing data can improve true mutation calls at the single cell level and filter out further noise.
In the future, some of this residual variation could be explained by individual medical histories or phenotypes, when they become available. In addition, as mentioned above, gene expression information can be extracted from the same cellular barcodes and from the very same genes whose SNVs have been used to generate the cell tree.

Efforts are currently underway to see how much of the currently unexplained variation can be accounted for by the cellular phenotypes and the individual medical histories. Translational geroscience seeks to identify which elements of the aging process are irreversible under current available treatments and which are amenable to modification by existing therapeutic interventions. The tree features involved in Cell Tree Rings may be valuable in diagnosing the long-term influence of a particular intervention by examining its effect on tree shape.

When extending Cell Tree Rings to other tissues an important question is how spatial aspects of the tree can be incorporated. Cellular elements in complex biofluids, such as blood and saliva, have considerable freedom of movement throughout the human body. In contrast, resident cells of compact, solid tissues in the kidney, intestine, liver for instance are under considerable spatial restrictions. The infiltrating immune cells of compact tissues are less constricted spatially than the resident cells, but more constricted than circulating blood cells. It is an open question how tree shape metrics contributing to the age regression model will change in a more restricted spatial environment. Spatial restrictions have been shown to be important in cancer where evolutionary phylodynamic models have been applied to model boundary-driven solid tumor growth [Lewinsohn et al, 2022] Combining spatial information with the temporal information of cell trees would thus help improve the ability of cell trees to quantify biological age.

Questions that Cell Tree Rings can help answer

One surprising finding in the phylogenetic tree-aided developmental biology literature is the degree of asymmetry in phylogenetic lineage trees [Bizzotto et al, 2021, Fasching et al, 2021]. These studies showed how, at least on the few samples studied, there can be a substantial difference in the number of surviving progeny between offspring of the first or first few cell divisions, the asymmetry reaching sometimes as large as 10:90%. Cell Tree Rings can help quantify this developmental imbalance further and separate it from changes in later life.

Somatic mutations in a small set of (growth and cancer associated) genes have been shown to propagate clones that become dominant in the hematopoietic system of older individuals [Mitchell et al, 2022] and have been directly linked to increased risk of cancer and other chronic diseases [Marongiu and DeGregori, 2022]. Additionally,

somatic mosaicism has been shown, in multiple tissues, to rise with age and to predict disease in animal models [Evans and Walsh, 2022]

From the perspective of designing interventions, it is important to understand which of the somatic mutations are simply passive indicators and which are active drivers of the aging process. In addition, it will be important to establish which can be targeted with clinical interventions.

By providing a way to quantify the different aspects of tree shape, Cell Tree Rings can be used to identify early indicators of clonal hematopoiesis and diagnose why certain individuals display resilience to the effects of somatic mutation and experience reduced chronic age-associated disease.

Cell Tree Rings and the timescales and convergence of different clocks

When evaluating the effectiveness of different biological aging clocks, it is important to address the question of what the minimal meaningful temporal unit of biological aging is. While clocks with high temporal resolution can evaluate the short-term effects of interventions, these effects can often be difficult to distinguish from physiological noise. On the other hand, lower temporal resolution over longer time windows may miss important short-term signals [Gabbutt et al, 2022]. Cell Tree Rings captures the long-term dynamics of somatic evolution that relates to the decades-long processes usually associated with aging but is insensitive to processes that are shorter than the characteristic timescale for detecting somatic mutations. It is an open question whether clocks based on different mechanisms and with different time resolutions can be combined and merged, but ultimately the results from different clocks should be reconciled.

Our hope is that Cell Tree Rings may provide a baseline integrative framework for different aging hallmarks and clocks. There are three reasons this may be possible:

First, Cell Tree Rings operate on the genome-instability level by tracking somatic mutations in hundreds or thousands of single cells. Importantly, it is the use of the tree structure to constrain these mutations that helps to improve their detection accuracy.

Second, Cell Tree Rings is based on a foundational construct, the somatic evolutionary cell tree, that relates the tens of trillions of somatic cells of a human body to each other and to time.

Third, without identifying the damage somatic mutations cause, it is difficult to design healthy longevity therapies and regimens. Cell Tree Rings captures and organizes this

basic mutation information at different levels of the tree hierarchy, potentially providing signposts for which interventions are likely to be most effective.

Cell Tree Rings is thus not simply another aging timer. It aims to provide a foundational principle for clocks. There is considerable uncertainty about whether epigenetic aging clocks can inform us about biological age reversals in clinical trials [Higgins-Chen et al, 2022]. Adding Cell Tree Rings as a single-cell resolution clock component might mitigate this uncertainty and improve the assessment of geroprotective trials.

Methods

Experimental data and protocol

Biological sample collection and isolation of cells: 18 blood samples, 5 ml each, have been collected by venipuncture at the Healthy Longevity Clinic in Prague, Czech Republic. The samples have been taken with informed consent from healthy patients of the clinic. The Healthy Longevity Clinic Ethical Committee has reviewed and approved the Tree Ring Pilot observational study protocol with the reference number 20220301_001. The age range of the volunteers was 21-82 years old at the time of blood collection, 10 volunteers were males and 8 were females. Samples were processed by the same protocol. In the following the data related to the first 6 samples are detailed. Viable peripheral blood mononuclear cells (PBMCs) were isolated from the collected biological sample. 4 ml of peripheral blood was diluted with 4 ml of 2% Fetal bovine serum (FBS) in Phosphate buffered saline (PBS). Subsequently, 8 ml of diluted peripheral blood was carefully layered on top of 4 ml of a density gradient (such as Lymphoprep™) and centrifuged at 300 g for 30 min. The cells were carefully harvested from the interface with a plastic pasteur pipette. Then, another 6 ml of 2% FBS/PBS is added to the cells, and then centrifuged at 300 g for 8 min with discarding of the supernatant and resuspending the cells in 1 ml of lysis solution. After one-minute incubation on ice, 4 ml of 2% FBS/PBS was added to the cells and centrifuged at 300g for 5 min with discarding of the supernatant and resuspending the cells in 1 ml of 2% FBS/PBS. Subsequently, the vitality and concentration of cells was determined through Acridine Orange and Propidium Iodide assay at LUNA Automated Cell

Counter. Cell concentration range was between $3.72x10^6$ – $6.35x10^6$ b/ml, and cell viability was between 99.1-99.7%.

Labeling the cells with CellPlex: The cells were labelled with molecular tags or CellPlex (according to original protocol CG000391 Cell Labeling with Cell Multiplexing Oligo RevA). Later, a specific volume of each sample was transferred into new 2 ml tubes and after labeling, the cells were washed 3 times with 2% FBS/PBS (instead of 2 times in comparison with the original protocol). After the last wash, the cells were resuspended in 600 µl of 2% FBS/PBS and counted at LUNA. Cell concentration range was between $3.15x10^6$ – $3.95x10^6$ b/ml, and cell viability was between 99.3-99.7%, post labeling.

The samples were pooled proportionally, and the final pool is passed through a 30 µm filter. Finally, the cells were counted and diluted to optimal concentration.

Loading and library preparation: Cells were loaded on to the Chromium Controller and libraries prepared according to original protocol CG000390 Chromium Next GEM Single Cell3 v3.1 Cell Surface Protein Cell Multiplexing RevB aiming for 16000 recovered cells. Some library preparation steps have been modified slightly. During cDNA amplification, the polymerisation step was extended to 1.5 min. After cDNA purification, the samples were split into two aliquots (A and B) that were processed in parallel and differed only in the implementation of size selection. After fragmentation, double size selection was modified for samples according to Table 3 below.

Step	**Sample**	**Volume of 1. SPRI [µl]**	**Transfer volume [µl]**	**Volume of 2. SPRI [µl]**
After fragmentation	A	20 (0.4x)	65	15 (0.7x)
	B	25 (0.5x)	70	5 (0.6x)
After PCR	A	50 (0.5x)	140	10 (0.6x)
	B	50 (0.5x)	140	10 (0.6x)

Table 1. Double-sided size selection using SPRIselect beads

After PCR amplification, both samples were purified using SPRIselect beads according to Table 1. At last, quality and quantity of libraries was determined using Fragment Analyzer and QuantiFluor dsDNA System.

The various chemicals or kits used were Next GEM Chip G Single Cell Kit, Next GEM Single Cell 3’ Gel Beads Kit v3.1, Next GEM Single Cell 3’ GEM Kit v3.1, Dynabeads MyOne Silane, Next GEM Single Cell 3’ Library Kit v3.1, Single Index Kit T Set A, 3’ CellPlex Kit Set A, 3’ Feature Barcode Kit, and Dual Index Kit NN SetA.

Sequencing: Library pools were sequenced on an Illumina NovaSeq 6000 using the S4 300-cycle kit and with 150 bp long R2.

Bioinformatics Processing of the experimental data

The output sequencing files have been processed with Cell Ranger v6.0.2 and the indexed paired end bam files have been converted into fastq files with bamtofastq v2.30.0. The fastq files were used for further processing.

Somatic mutation calling de novo from scRNA-seq

We have used scSNV v1.0b to call somatic mutations, specifically single nucleotide variants, directly from 10X Genomics scRNA-seq data [Wilson et al, 2021] through collapsed molecular duplicates to increase mutation coverage. GRCh37 (hg19) reference human genome build was used for mapping and alignment. The default settings have been used except the variant allele fraction was set to 0.75. The input were fastq files and the output were matrices containing allele specific alternative and reference counts for each SNVs, SNV and cell count matrices and VCF files.

Phylogenetic Tree Inference

The matrices and VCF files generated in the previous step are used to generate fasta alignments. Fasta files are generated from all the cells and a subset of the cells with SeqKit v2.1.0 [Shen et al, 2016]. The trees are generated using 25 different pseudo-replicates, where a pseudo-replicate is SNV mutation information called from 700 separate cells per individual sample. It is a pseudo-replicate since the 700 cells do partially overlap between different replicates. In order to decrease the number of false positives we required for one alternative allele variant to be detected by at least 11 molecules with different UMIs per cell.

In terms of phylogenetics inference with UPGMA, the R package phangorn v2.8.1 [Schliep, 2010] has been used with helper functions from the ape package v5.6.2. The substitution model used was Felsenstein's F81, and the matrix of pairwise distances was computed with the dist.dna function of the ape package. Tree inference provided rooted, ultrametric trees by default. The trees were stored in newick files.

Cell Trees have been visualised with version 1.4.4 of the FigTree tree figure drawing tool.

Cell Tree Metrics

The features can be split into 5 groups based on their technical properties. Group I contains spectral tree metrics that are based on the so called transform matrices of the cell trees as the discrete analogues of the Fourier [Hicks et al, 2019] and Laplacian transforms, correspondingly. Group II contains specialised phylogenetic features focusing on aggregated branch length statistics and their derivatives, like entropy based metrics. Group III includes well-known general phylogenetic tree statistics used in the biodiversity. Group IV is focusing on branch length values specifically and generates summary statistics based on the distance matrix between the tips of the tree. Finally, Group V has 2 powergraph based features generating first the Laplacian transforms of the square of the tree graphs, similarly to Group I.

Statistical Analysis

Statistical analysis was performed using Python (version 3.9.7). The scikit-learn module sklearn.linear_model.LassoLarsCV was used for cross-validation and model selection with LARS-Lasso. LARS, which stands for Least Angle Regression, is a computationally efficient way to cross-validate the lasso [Efron et al, 2004]. Leave-one-out cross-validation was used to estimate the regularisation parameter providing the minimum RMSE (root mean square) error. Interaction terms have been used between the predictor variables.

References

Bizzotto S, Dou Y, Ganz J, Doan RN, Kwon M, Bohrson CL, Kim SN, Bae T, Abyzov A; NIMH Brain Somatic Mosaicism Network, Park PJ, Walsh CA. Landmarks of human embryonic development inscribed in somatic mutations. Science. 2021 Mar 19;371(6535):1249-1253. doi: 10.1126/science.abe1544. PMID: 33737485; PMCID: PMC8170505.

Csordas A. Cell lineage trees: the central structure plus key dynamics of biological aging and formulating the limiting problem of comprehensive organismal rejuvenation PeerJ Preprints, v1, 24 June 2019

Efron B., Hastie T., Johnstone Iain, Tibshirani R. Least Angle Regression. The Annals of Statistics 2004, Vol. 32, No. 2, 407–451 DOI: 10.1214/009053604000000067

Evans MA, Walsh K. Clonal hematopoiesis, somatic mosaicism, and age-associated disease. Physiol Rev. 2023 Jan 1;103(1):649-716. doi: 10.1152/physrev.00004.2022. Epub 2022 Sep 1. PMID: 36049115; PMCID: PMC9639777.

Fasching L, Jang Y, Tomasi S, Schreiner J, Tomasini L, Brady MV, Bae T, Sarangi V, Vasmatzis N, Wang Y, Szekely A, Fernandez TV, Leckman JF, Abyzov A, Vaccarino FM. Early developmental asymmetries in cell lineage trees in living individuals. Science. 2021 Mar 19;371(6535):1245-1248. doi: 10.1126/science.abe0981. PMID: 33737484; PMCID: PMC8324008.

Gabbutt C, Schenck RO, Weisenberger DJ, Kimberley C, Berner A, Household J, Lakatos E, Robertson-Tessi M, Martin I, Patel R, Clark SK, Latchford A, Barnes CP, Leedham SJ, Anderson ARA, Graham TA, Shibata D. Fluctuating methylation clocks for cell lineage tracing at high temporal resolution in human tissues. Nat Biotechnol. 2022 May;40(5):720-730. doi: 10.1038/s41587-021-01109-w. Epub 2022 Jan 3. PMID: 34980912; PMCID: PMC9110299.

Hastie T, Tibshirani R, Wainwright M. Statistical Learning with Sparsity: The Lasso and Generalizations, CRC Press, 2015

Hicks DG, Speed TP, Yassin M, Russell SM. Maps of variability in cell lineage trees. PLoS Comput Biol. 2019 Feb 12;15(2):e1006745. doi: 10.1371/journal.pcbi.1006745. PMID: 30753182; PMCID: PMC6388934.

Higgins-Chen AT, Thrush KL, Wang Y, Minteer CJ, Kuo PL, Wang M, Niimi P, Sturm G, Lin J, Moore AZ, Bandinelli S, Vinkers CH, Vermetten E, Rutten BPF, Geuze E, Okhuijsen-Pfeifer C, van der Horst MZ, Schreiter S, Gutwinski S, Luykx JJ, Picard M, Ferrucci L, Crimmins EM, Boks MP, Hägg S, Hu-Seliger TT, Levine ME. A computational solution for bolstering reliability of epigenetic clocks: Implications for clinical trials and longitudinal tracking. Nat Aging. 2022 Jul;2(7):644-661. doi: 10.1038/s43587-022-00248-2. Epub 2022 Jul 15. PMID: 36277076; PMCID: PMC9586209.

Kaeberlein M. Translational geroscience: A new paradigm for 21st century medicine. Transl Med Aging. 2017 Oct;1:1-4. doi: 10.1016/j.tma.2017.09.004. Epub 2017 Sep 27. PMID: 32219192; PMCID: PMC7098696.

Lähnemann D, Köster J, Szczurek E, McCarthy DJ, Hicks SC, Robinson MD, Vallejos CA, Campbell KR, Beerenwinkel N, Mahfouz A, Pinello L, Skums P, Stamatakis A, Attolini CS, Aparicio S, Baaijens J, Balvert M, Barbanson B, Cappuccio A, Corleone G, Dutilh BE, Florescu M, Guryev V, Holmer R, Jahn K, Lobo TJ, Keizer EM, Khatri I, Kielbasa SM, Korbel JO, Kozlov AM, Kuo TH, Lelieveldt BPF, Mandoiu II, Marioni JC, Marschall T, Mölder F, Niknejad A, Raczkowski L, Reinders M, Ridder J, Saliba AE, Somarakis A, Stegle O, Theis FJ, Yang H, Zelikovsky A, McHardy AC, Raphael BJ, Shah SP, Schönhuth A. Eleven grand challenges in single-cell data science. Genome Biol. 2020 Feb 7;21(1):31. doi: 10.1186/s13059-020-1926-6. PMID: 32033589; PMCID: PMC7007675.

Lewinsohn MA, Bedford T, Müller NF.,Alison FF. State-dependent evolutionary models reveal modes of solid tumor growth. 2022, doi:10.1101/2022.08.05.502978 biorxiv.org

Lewitus E, Morlon H. Characterizing and Comparing Phylogenies from their Laplacian Spectrum. Syst Biol. 2016 May;65(3):495-507. doi: 10.1093/sysbio/syv116. Epub 2015 Dec 12. Erratum in: Syst Biol. 2016 Sep;65(5):944. PMID: 26658901.

Lodato MA, Rodin RE, Bohrson CL, Coulter ME, Barton AR, Kwon M, Sherman MA, Vitzthum CM, Luquette LJ, Yandava CN, Yang P, Chittenden TW, Hatem NE, Ryu SC, Woodworth MB, Park PJ, Walsh CA (2018) Aging and neurodegeneration are associated with increased mutations in single human neurons. Science 359(6375):555–559

López-Otín C, Blasco MA, Partridge L, Serrano M, Kroemer G. The hallmarks of aging. Cell. 2013 Jun 6;153(6):1194-217. doi: 10.1016/j.cell.2013.05.039. PMID: 23746838; PMCID: PMC3836174.

Macdonald-Dunlop E, Taba N, Klarić L, Frkatović A, Walker R, Hayward C, Esko T, Haley C, Fischer K, Wilson JF, Joshi PK. A catalogue of omics biological ageing clocks reveals substantial commonality and associations with disease risk. Aging (Albany NY). 2022 Jan 24;14(2):623-659. Doi: 10.18632/aging.203847. Epub 2022 Jan 24.

Marongiu F, DeGregori J. The sculpting of somatic mutational landscapes by evolutionary forces and their impacts on aging-related disease. Mol Oncol. 2022 Sep;16(18):3238-3258. doi: 10.1002/1878-0261.13275. Epub 2022 Jun 29. PMID: 35726685; PMCID: PMC9490148.

Mitchell E, Spencer Chapman M, Williams N, Dawson KJ, Mende N, Calderbank EF, Jung H, Mitchell T, Coorens THH, Spencer DH, Machado H, Lee-Six H, Davies M, Hayler D, Fabre MA, Mahbubani K, Abascal F, Cagan A, Vassiliou GS, Baxter J, Martincorena I, Stratton MR, Kent DG, Chatterjee K, Parsy KS, Green AR, Nangalia J, Laurenti E, Campbell PJ. Clonal dynamics of haematopoiesis across the human lifespan. Nature. 2022 Jun;606(7913):343-350. doi: 10.1038/s41586-022-04786-y. Epub 2022 Jun 1. PMID: 35650442; PMCID: PMC9177428.

Partridge L, Deelen J, Slagboom PE. Facing up to the global challenges of ageing. Nature. 2018 Sep;561(7721):45-56. doi: 10.1038/s41586-018-0457-8. Epub 2018 Sep 5. PMID: 30185958.

Rutledge, J., Oh, H. & Wyss-Coray, T. Measuring biological age using omics data. *Nat Rev Genet* **23**, 715–727 (2022). https://doi.org/10.1038/s41576-022-00511-7

Salipante SJ, Horwitz MS. Phylogenetic fate mapping. Proc Natl Acad Sci U S A. 2006 Apr 4;103(14):5448-53. doi: 10.1073/pnas.0601265103. Epub 2006 Mar 28. PMID: 16569691; PMCID: PMC1414797.

Sankaran VG, Weissman JS, Zon LI. Cellular barcoding to decipher clonal dynamics in disease. Science. 2022 Oct 14;378(6616):eabm5874. doi: 10.1126/science.abm5874. Epub 2022 Oct 14. PMID: 36227997.

Schliep KP. phangorn: phylogenetic analysis in R. Bioinformatics. 2011 Feb 15;27(4):592-3. doi: 10.1093/bioinformatics/btq706. Epub 2010 Dec 17. PMID: 21169378; PMCID: PMC3035803.

Sender R, Fuchs S, Milo R. Revised Estimates for the Number of Human and Bacteria Cells in the Body. PLoS Biol. 2016 Aug 19;14(8):e1002533. doi: 10.1371/journal.pbio.1002533. PMID: 27541692; PMCID: PMC4991899.

Shen W, Le S, Li Y, Hu F. SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation. PLoS One. 2016 Oct 5;11(10):e0163962. doi: 10.1371/journal.pone.0163962. PMID: 27706213; PMCID: PMC5051824.

Stadler T, Pybus OG, Stumpf MPH. Phylodynamics for cell biologists. Science. 2021 Jan 15;371(6526):eaah6266. doi: 10.1126/science.aah6266. PMID: 33446527.

Wasserstrom A, Frumkin D, Adar R, Itzkovitz S, Stern T, Kaplan S, Shefer G, Shur I, Zangi L, Reizel Y, Harmelin A, Dor Y, Dekel N, Reisner Y, Benayahu D, Tzahor E, Segal E, Shapiro E. Estimating cell depth from somatic mutations. PLoS Comput Biol. 2008 May 9;4(4):e1000058. doi: 10.1371/journal.pcbi.1000058. PMID: 18404205; PMCID: PMC2275312.

Wilson GW, Derouet M, Darling GE, Yeung JC. scSNV: accurate dscRNA-seq SNV co-expression analysis using duplicate tag collapsing. Genome Biol. 2021 May 7;22(1):144. doi: 10.1186/s13059-021-02364-5. PMID: 33962667; PMCID: PMC8103760.

Contributions

A.Cs. conceived the project and wrote the original manuscript. D.G.H. and B.S. edited the manuscript. B.S. produced the core software pipeline. A.Cs., B.S. and D.G.H. designed the study, the methodology and wrote software. A.Cs. and D.G.H. performed the statistical analysis and supervised the study. A.V collected the blood and F.Z supervised the clinical procedure. T.K and B.T. performed library preparation and scRNA-seq. T.K. wrote the original draft of the Experimental protocol. All authors read and approved the final version of the manuscript.

Corresponding author

Attila Csordas, email: attila@agecurve.co.uk.

Acknowledgements

This work was solely supported by AgeCurve Limited. Special acknowledgement goes to Petr Sramek, of LongevityTech.Fund for providing crucial infrastructure in the Czech Republic. We would also like to acknowledge the patients of Healthy Longevity Clinic, who volunteered to provide blood. Special thanks to Gavin Wilson for insights on the scSNV pipeline, and Kylie Chen for phylogenetics consultation.

Competing interests

AgeCurve Limited has filed a patent called Cell Tree Rings: method and cell lineage tree based aging timer for calculating biological age of a biological sample. A.Cs. is a shareholder, D.G.H. and B.S. are option holders of AgeCurve Limited.