

**Sejtfa-életkor mint új evolúciós  
modell az életkorral összefüggő  
szomatikus mutációs teher  
reprezentálására**

**Tézisfüzet**

**Csordás Attila**  
Doktori értekezés

Szeged  
2024

**Sejtfa-életkor mint új evolúciós modell az  
életkorral összefüggő szomatikus mutációs  
teher ábrázolására**



**Csordás Attila**  
**Doktori értekezés**

Megbízott konzulens: Prof. Dr. Kemény Lajos

Szegedi Tudományegyetem  
Szent-Györgyi Albert Orvostudományi Kar  
Klinikai Orvostudományi Doktori Iskola  
Szeged  
2024

## Összefoglalás

A biológiai életkor becslése általában olyan biomarkerek segítségével történik, amelyek állapota megfigyelések szerint korrelál a kronológiai életkorral. Az ilyen öregedési órák egyik gyengesége, hogy nehéz megállapítani, hogy a biomarker állapotok hogyan kapcsolódnak az öregedés mechanizmusaihoz. A szomatikus mutációk potenciálisan egy alapvetőbb öregedési óra alapját képezhetik, mivel a mutációk az öregedés markerei és mozgatórugói is egyben, és természetes időskálával rendelkeznek. Az ilyen mutációkból épített sejtvonalfák tükrözik a szomatikus evolúciós folyamatot, és így - feltételezések szerint - a szervezet öregedési állapotát. Egy ilyen öregedési időzítő azonban eddig nem volt kivitelezhető, mivel a szomatikus variánsok kimutatása az egyes sejtekben jelentős technológiai kihívást jelent.

Eredményeink szerint az egysejtes RNS-szekvenálás (scRNA-seq) segítségével több ezer sejtből detektált szomatikus mutációk felhasználhatók egy olyan sejtvonalfára felépítésére, amelynek szerkezete korrelál a

kronológiai korról. De novo egyedi nukleotid-variánsokat (SNV-k) detektálunk emberi perifériás vér mononukleáris sejtekben egy módosított protokoll segítségével. A kronológiai életkoron regresszált 31 filogenetikai sejtfa prediktort felhasználó, többszörös regresszióján alapuló alapértelmezett modell 0,81-es Pearson-korrelációt és ~4 év abszolút hiba mediánját adja a megjósolt és a kronológiai életkor között. A modell tesztelése egy nyilvános scRNA-seq-adathalmazon 0,85-ös Pearson-korrelációt eredményez. Ezen túlmenően a sejtfa-életkor előrejelzései bizonyos klinikai biomarkerek, például a glükóz, az albuminszint és a leukocitaszám esetében jobb prediktornak bizonyultak, mint a kronológiai életkor önmagában.

A sejtvonalfa geometriája rögzíti az egyén szomatikus evolúciójának szerkezetét, és egy új öregedési időzítőtípust képvisel. Amellett, hogy numerikus becslést ad a sejtfa-életkorról, feltárja az öregedési folyamat időbeli történetét, megmutatva, hogyan fejlődik a klonális struktúra az élettartam során. A Sejtfagyűrűk (Cell Tree Rings) modell kiegészíti a meglévő öregedési órákat, és

segíthet csökkenteni a geroprotektív kísérletek kiértékelésének jelenlegi bizonytalanságát.

Az emberben felhalmozódó szomatikus mutációs teher kódolása evolúciós sejtfaikon keresztül, és az általános szomatikus mutációs teher numerikus ábrázolása sejtfá-életkorként új modellt nyújt az öregedés legalapvetőbb jellemzőjének összegzésére, a biológia talán legjobban bevált kvantitatív módszertanának, a filogenetikának a felhasználásával.

## Kutatási célkitűzések

Az általunk javasolt öregedési időmérő központi feltételezése az, hogy a sejtfaik szerkezete vagy "alakja" a biológiai öregedési folyamatot reprezentálja. Ennek a hipotézisnek két oka van. Az első az, hogy a filogenetikai szisztematika már régóta kimutatta, hogy a ma létező fajok közötti genetikai távolságok hogyan tükrözik a múltban bekövetkezett evolúciós változásokat. Ésszerű tehát azt várni, hogy az egyes sejtek közötti genetikai távolságok felhasználhatók a sejtek szomatikus evolúciós történetére való következtetésre, ami az öregedés egyik mozgatórugója és indikátora. A második az, hogy az

orvosbiológiai élettörténet nyomot hagyhat a sejtfaán, és így az öregedési folyamat főbb átmeneteiről is képet adhat. A sejtfaák további előnye, hogy az öregedés dinamikájának intuitívan vonzó ábrázolását nyújtják, amely segíthet a további interpretációban.

Egészséges egyének (n=18, 21-82 éves korosztály) emberi perifériás vérsejtjeinek felhasználásával kifejlesztettünk egy új öregedési időmérőt, a Sejtfa-gyűrűket (CTR), amely a következő jellemzőkkel rendelkezik:

1. A természetesen előforduló szomatikus egyedi nukleotid-variánsokat (SNV-k) használjuk a sejtfaák felépítéséhez, standard filogenetikai algoritmusok segítségével,
2. Az SNV-ket közvetlenül és de novo hívják meg több száz vagy több ezer sejtéből származó scRNA-seq adatokból,
3. A fa metrikák széles skáláját használjuk a fa alakjának azon aspektusainak azonosítására, amelyek egy többszörös regressziós modell segítségével összefüggésbe hozhatók a kronológiai korrall,

4. A modellt arra használjuk, hogy megjósoljuk az egyének sejtfá-életkorát.

Két különböző típusú filogenetikai algoritmus, az UPGMA és a maximum likelihood, működőképes sejtfá-életkor modellt állít elő, ami bizonyítékot szolgáltat a hipotézishez. A modellt független tesztkészletként nyilvános adatokkal validáljuk. A prediktált sejtfá-életkorok korrelálnak néhány klinikai vér biomarkerrel, például a glükózzal, az albuminnal, a leukocitákkal és a monocitákkal.

## Módszerek

### Kísérleti adatok és vizsgálati protokoll

Biológiai mintagyűjtés és sejtek izolálása: 18 darab, egyenként 5 ml-es vérmintát vénapunkcióval gyűjtöttek a prágai Healthy Longevity Klinikán (HLC), Csehországban. A mintákat a klinika egészséges pácienseitől tájékozott beleegyezéssel vették. A Healthy Longevity Klinika etikai bizottsága felülvizsgálta és jóváhagyta a Tree Ring Pilot megfigyelési tanulmány protokollját a 20220301\_001 hivatkozási számmal. Az

önkéntesek életkora a vérvétel időpontjában 21-82 év volt, 10 önkéntes férfi és 8 nő volt. A minták feldolgozása azonos protokoll szerint történt. A gyűjtött biológiai mintából életképes perifériás vér mononukleáris sejteket (PBMC) izoláltak.

A sejteket molekuláris címkével vagy CellPlexszel jelölték.

A sejteket a Chromium Controllerre töltöttük, és könyvtárakat készítettünk az eredeti CG000390 Chromium Next GEM Single Cell<sup>3</sup> v3.1 Cell Surface Protein Cell Multiplexing RevB protokoll szerint, 16000 visszanyert sejtre törekedve. Néhány könyvtárkészítési lépést kissé módosítottunk. A cDNS-amplifikáció során a polimerizációs lépést 1,5 percre hosszabbítottuk meg. A cDNS-tisztítás után a mintákat két aliquotra osztottuk (A és B), amelyeket párhuzamosan dolgoztunk fel, és csak a méretszelekció végrehajtásában különböztek.

A PCR-amplifikációt követően mindkét mintát SPRIselect gyöngyökkel tisztítottuk.

Szekvenálás: A könyvtár poolokat Illumina NovaSeq 6000-en szekvenáltuk az S4 300 ciklusú kit használatával és 150 bp hosszú R2-vel.



## A Sejtfa-gyűrűk modell lépései

1. Tizkit: SNV-k vonalkódspecifikus meghívása scSNV-vel és csírvonal-szűréssel.
2. Tiznit: Fasta fájlok generálása és sejtfa filogenetikai építése.
3. AgeTreeShape: Fa statisztikai jellemzőinek kiszámítása és egyváltozós regresszió az életkorra.
4. CellTreeAge: Többszörös regressziós modell építése és sejtfa-életkor előrejelzése.

## Eredmények

### Sejtfa metrikák

Az értekezés központi hipotézise az, hogy a sejtfa alakja a biológiai életkor mérőszáma. Az alak itt a topológia (elágazási sorrend) és az ághossz kombinációjára utal. A topológia a sejtvonalfák esetében a szomatikus sejtek mitotikus elágazási mintázatának felel meg. Az ághossz az evolúciós változás mértékét jelenti, és általában a

mutációs ráták és egy alkalmas időegység szorzataként határozzák meg.

A különböző fametrikák vagy csak a topológiát, vagy csak az ághosszat, vagy a kettő kombinációját mérik. Egy 31 fametrikából álló készletet alkalmaztunk emberi perifériás vér mononukleáris sejtek szomatikus mutációiból felépített sejtfa alakjának jellemzésére. Ez a mérőszámkészlet magában foglalja a filogenetikában használt hagyományos fa metrikákat és néhány olyan mérőszámot is, amelyeket kifejezetten ehhez a tanulmányhoz fejlesztettünk ki.

A fametrikák, vagy jellemzők technikai tulajdonságaik alapján 5 csoportra oszthatók. Az I. csoport a spektrális fametrikákat tartalmazza, amelyek a sejtfa transzformációs mátrixain alapulnak, és az általánosított Fourier-, illetve Laplace-transzformáció diszkrét analógjai. A II. csoport speciális filogenetikai jellemzőket tartalmaz, amelyek az aggregált ághossz-statisztikákra és azok származékaira összpontosítanak, beleértve az entrópia-alapú metrikákat is. A III. csoport a biológiai diverzitással kapcsolatos szakirodalomban használt, jól ismert általános filogenetikai fa-statisztikákat

tartalmazza. A IV. csoport kifejezetten az ághosszértékekre összpontosít, és a fa csúcsai közötti távolságmátrixon alapuló összefoglaló statisztikákat készít. Végül az V. csoport az I. csoporthoz hasonlóan 2 powergraph-alapú jellemzőt tartalmaz, amelyek a fa gráfok négyzetének Laplace-transzformációit generálják.

### Alapértelmezett modellépítés és előrejelzési eredmények az összes fametrika felhasználásával

Az sejtfa-életkor előrejelzőt úgy állítottuk elő, hogy a kronológiai életkort regresszáltuk a 31 fametrikára, valamint a biológiai nemre. A modell validálására és tesztelésére Elastic Net regularizációt és 'nested leave-one-out' keresztvalidálást használtunk. Az így kapott előrejelzési hibákat, korrelációs együtthatót, p-értékeket és megmagyarázott varianciát a prediktált sejtfa-életkorok és a kronológiai életkor összehasonlításával becsültük meg.

Az alapértelmezett modellben a filogenetikai fa következtetéshez UPGMA-t alkalmaztunk, mintánként 5

pszeudo-replikátum fát használtunk, és mind a 32 jellemzőt felhasználtuk a regresszióban. Ennél a modellnél  $r=0,813$ ,  $p=0,00004$ ,  $R^2=0,660$ ,  $MAE=7,625$ ,  $MdAE=4,396$  és  $RMSE=10,760$ .

### A sejtfá-életkor modell validálása publikus adatokon

Miután a sejtfá-életkor modellt HLC-adatokon kifejlesztettük, teszteltük azt egy nyilvános scRNA-seq adatkészleten. Ehhez az Asian Immune Diversity Atlas (AIDA) 18 perifériás vérmintáját használtuk. Ebben a mintában 10 nő és 8 férfi volt, életkoruk 21-65 év között mozgott. Ezeknél az adatoknál a HLC-adatokhoz hasonlóan UPGMA-t használtunk, hogy minden mintából 5 pszeudo-replikátum fát hozzunk létre, ahol minden pszeudo-replikátum 700, a mintából véletlenszerűen kiválasztott sejtet használt. Az alapértelmezett modellt mind a 18 HLC-adaton betanítottuk. Az eredményül kapott teljesítménymutatók a következők voltak:  $r=0,853$ ,  $p=0,00001$ ,  $R^2=0,728$ ,  $MAE=12,791$ ,  $MdAE=13,636$ ,  $RMSE=14,081$ .

## Sejtfa-életkor asszociációk klinikai biomarkerekkel

A jóváhagyott vizsgálati protokoll részeként a klinikai vér biomarkerek széles skálája állt rendelkezésre. Ezeket ugyanabban az időszakban nyerték, mint az scRNA-seq vizsgálathoz szükséges vérvételeket. A klinikai vérmarkerek alábbi listájához a 18 HLC-mintából 13 minta értékei álltak rendelkezésünkre, 9 férfi és 4 nő esetében: teljes leukocita és eritrocita vérpanel (13 marker), albumin, glükóz, hba1c, összkoleszterin, alkalikus foszfatáz, húgysav, kreatinin, vér karbamid nitrogén. Ez összesen 21 markert jelentett.

A 21 vérmarker közül 8 marker szignifikánsan összefüggött prediktált sejtfa-életkorral vagy a kronológiai korrall, vagy mindkettővel. E 8 szignifikáns összefüggésből hat esetében a sejtfa-életkor erősebb előrejelzőnek bizonyult, mint a kronológiai kor, a legnagyobb különbséget a vércukor metabolikus marker mutatta.

## A dolgozat alapját képező publikációk listája

- I. **Csordas A**, Sipos B, Kurucova T, Volfova A, Zamola F, Tichy B, Hicks DG. Cell Tree Rings: the structure of somatic evolution as a human aging timer. *Geroscience*. 2024 Jan 4. doi: 10.1007/s11357-023-01053-4. Epub ahead of print. PMID: 38172489. **Q1 with IF:5.7**
- II. **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* (2019). **Perspective, not peer reviewed.** <https://peerj.com/preprints/27821v7/>