

A MAGYAR  
TUDOMÁNY  
ÜNNEPE

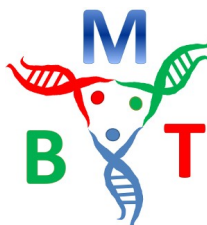


Magyar Tudományos  
Akadémia



# Bioinformatika 2023

A Magyar Tudományos Akadémia Bioinformatikai Osztályközi Tudományos Bizottsága  
és a Magyar Bioinformatikai Társaság tudományos konferenciája



**2023. november 10.**

# Program

10:00 - 10:05: **Köszöntő (Prof. Pongor Sándor)**

## 1. szekció - Bioinformatikai algoritmusok

**Elnök: Dr. Dosztányi Zsuzsanna**

10:05 - 10:25: **Szöllősi J Gergely**

Koevolúciós függőségek biológiai skálákon átívelő modellezése  
*MTA-ELTE „Lendület” Evolúciós Genomika Kutatócsoport*

10:25 - 10:45: **Ligeti Balázs**

Genomikai nyelvmodellek a mikrobiomkutatásban  
*PKKE Információs Technológiai és Bionikai Kar*

10:45 - 11:00: **Erdős Gábor**

AIUPred: új mély tanulós módszer rendezetlen régiók predikciójára  
*MTA-ELTE Lendület Bioinformatikai Kutatócsoport*

11:00 - 11:15: **Németh Eszter**

Mutációs mintázatok: biológia vs. informatika  
*HUN-REN TTK Enzimológiai Intézet, Genomstabilitás Kutatócsoport*

11:15 - 11:30: **Pethő Gergely**

Transzformer és rekurrens architektúrát egyesítő hibrid neurális hálózat használata bakteriofágok és plazmidok detektálására  
*DE, Egy Egészség Intézet*

11:30 - 11:45: **Szünet**

## 2. szekció - Szerkezeti Bioinformatika

**Elnök: Prof. Gáspári Zoltán**

11:45 - 12:05: **Hetényi Csaba**

A hidratáció szerepe a gyógyszertervezésben  
*PTE Farmakológiai és Farmakoterápiai Intézet, PTE Farmakoinformatikai Tanszék*

12:05 - 12:25: **Bajusz Dávid**

SH2db: online fehérjeszerkezeti adatbázis az SH2 doménekre  
*HUN-REN TTK Szerves Kémiai Intézet, Gyógyszerkémiai Kutatócsoport*

12:25 - 12:40: **Lukács András**

Fehérjék másodlagos szerkezetének jóslása - a szürke zóna  
*ELTE TTK Matematikai Intézet, Mesterséges Intelligencia Kutatócsoport*

12:40 - 13:20: **Szünet**

## 3. szekció - Rendszerbiológia

**Elnök: Prof. Csikász-Nagy Attila**

13:10 - 13:40: **Papp Balázs**

A metabolom evolúciójától a humán betegségekig  
*HUN-REN SzBK Biokémiai Intézet, HCEMM-BRC*

13:40 - 14:00: **Szalai Bence**

Tumorsejtek gyógyszerérzékenységének predikciója - avagy hogyan vezethetnek félre az MI kiértékelési metrikák  
*Turbine Kft*

14:00 - 14:15: **Mikulecz Benjámín Áron**

A hibaküszöb egy egyszerű, kompartmentalizált rendszerben  
*ELTE, Biológia Doktori Iskola*

14:15 - 14:30: **Zakar-Polyák Enikő**

*Sejtek transzkriptomikai korának meghatározása öregedési órák segítségével*  
*HUN-REN SZTAKI*

14:30 - 14:45: **Fichó Erzsébet**

Hogyan segíthet a számítógépes sejszimuláció a hatékonyabb  
gyógyszerfejlesztésben?

*Cytocast Hungary Kft*

14:45 - 15:05: **Szünet**

#### **4. szekció - Orvosi bioinformatika**

**Elnök: Prof. Gyórfy Balázs**

15:05 - 15:25: **Ágg Bence**

Transzkriptomiai és fehérje-fehérje kölcsönhatási hálózatelemzés  
gyógyszercélpontok azonosítására

*SE ÁOK*

15:25 - 15:45: **Pongor S. Lőrinc**

A kissejtes tüdőrák genomikai jellemzése

*HCEMM Onkogenomikai és Epigenetikai Kutatócsoport*

15:45 - 16:00: **Dobolyi Zsófia**

Vesetumороk célzott terápiás célpontjainak kifejeződése egysejt RNS-  
szekvenálási adatok alapján

*SE, Bioinformatika Tanszék; HUN-REN, TTK, Onkológiai Biomarker Kutatócsoport*

16:00 - 16:15: **Fejes Iván**

Arc fotó alapú biológiai óra elemzése mortalitásra és foglalkozásokra

*HUN-REN Sztaki*

16:15 - 16:30: **Prechl József**

Kvantitatív szerológia növekedési függvények alkalmazásával

*Diagnosticum Zrt., K+F Laboratórium*

16:30 - 16:35: **Zárszó (Gyórfy Balázs)**

# Előadások kivonatai

## 1. szekció - Bioinformatikai algoritmusok

### Koevolúciós függőségek biológiai skálákon átívelő modellezése

Szólósi J Gergely

MTA-ELTE „Lendület” Evolúciós Genomika Kutatócsoport

A szekvencia- és genomevolúció jelenlegi legmodernebb modelljei az egyes szervezeti szinteken belüli függetlenség feltételezésén alapulnak. Noha ez a feltételezés nélkülözhetetlen a nagyméretű és folyamatosan növekvő szekvenciaadatok elemzéséhez, korlátozza az evolúciós múlt rekonstrukcióját és annak megértését, hogy a biológiai rendszerek különböző elemei, mint az aminosavak, gének vagy fajok, hogyan hatnak egymásra. Kutatásunk célja, hogy megszabaduljunk ezektől a korlátoktól. Hierarchikus probabilisztikus modellekkel és újabban machine learning eljárásokkal segítjük a jelenleg függetlennek kezelt koevolúciós függőségek modellezését.

Ezen modellek alkalmazásával meghatároztuk az baktériumok filogenetikai fájának gyökerét (Coleman et al. 2021), ami lehetővé tette a bakteriális közös ős rekonstrukcióját. Ezt követően a többsejtűség kialakulását vizsgálva az Opisthokonta szupergrupon belül az állatok és gombák genetikai változásainak útvonalát rekonstruáltuk (Ocaña-Pallarès et al. 2022) és megállapítottuk, hogy az állatok a többsejtűséghez szükséges gének felhalmozódása után jöttek létre, míg a gombák ősei a funkcionális kategóriák többségét elvesztették. A bakteriális fa mentén rekonstruált géntartalom-perdikciókat gépi tanulási módszerekkel ötvözve fel tudtuk használni a nagy Oxidációs Eseményt (Great Oxidation Event, GOE) arra, hogy időben kalibráljuk a bakteriális családfát.

### Genomikai nyelvmodellek a mikrobiomkutatásban

Ligeti Balázs, Bodnár Babett, Szepesi-Nagy István

PKKE Információs Technológiai és Bionikai Kar

A kvantitatív biológia egyik alapvető kérdése, hogy miként fedezhetünk fel újszerű mintázatokat és struktúrákat a biológiai adatokban, ami kulcsfontosságú az olyan összetett komplex szerveződések modellezéséhez, előrejelzéséhez és manipulálásához, mint a mikrobiom. A mai kihívások – például a mikrobiomok speciális, célzott feladatokra való alkalmazása – újfajta eljárások kidolgozását teszik szükségessé.

Kidolgoztunk egy olyan univerzálisan felhasználható szekvenciareprezentációt - a prokBERT-et -, ami lehetővé teszi a mikrobiomosztályozási feladatok hatékony megoldását az ún. transfer learning segítségével, illetve egy új, egyszerű vektorizációs eljárást - LCA (local context-aware) tokenization -, ami lehetővé teszi a transzformer alapú modellek hatékony tanítását. A modellt megközelítőleg 18000 assembly (baktérium, vírus, gomba és archaea szekvencia) alapján tanítottuk, ami mintegy 200 milliárd tokent jelent. Megmutattuk, hogy a csupán az ún. maszkolási feladaton, self-supervised módon előtanított modellel is hatékonyan elkülöníthetők a szekvenciaregiók - mint a kódozó (CDS) vagy intergenikus régiók -, ami alátámasztja a modell jó általánosító képességét. A ProkBERT modelleket emellett alkalmaztuk több osztályozási feladatban is, mint pl. promóterazonosítás és fág-baktérium szekvencia felismerése; mind a hagyományos (pl. Virsorter2), mind az alternatív deep learning (pl. INHERIT) eszközöknél jobban teljesítettek a modelljeink. További előnyük, hogy gyorsak (~5Mb/sec) és könnyen felhasználhatóak. A ProkBERT modellek nyilvánosan elérhetőek a github-on, illetve a Huggingface-en.

### AIUPred: új mély tanulós módszer rendezetlen régiók predikciójára

Erdős Gábor, Dosztányi Zsuzsanna

MTA-ELTE Lendület Bioinformatikai Kutatócsoport

A rendezetlen fehérjék kritikus szerepet játszanak különböző szubcelluláris folyamatokban, beleértve a szabályozást, a transzkripciót és az apoptózist. Elsősorban más fehérjék molekuláris felismerésében vesznek részt, ami gyakran egy kapcsolt hajtogatási és kötődési folyamaton keresztül vezet átmenethez egy rendezettebb állapotba. A funkcionális

rendezetlen régiókban bekövetkező mutációkat számos betegséggel hozták összefüggésbe, ami hangsúlyozza a rendezetlen régiók pontos jellemzésének szükségességét. A rendezetlenség vizsgálatára szolgáló kísérleti technikák kihívást jelentenek, költségesek és időigényesek, így egyre nagyobb szükség van a számítási eszközökre.

Kutatásunk során egy újszerű energia beágyazó transzformer hálózatot készítettünk, amely a globuláris fehérjék szerkezeti jellemzőinek magas dimenziós leírását hozza létre. Ezt a beágyazást a Protein Data Bankból származó nagyfelbontású monomer struktúrákon tanítottuk be, egy Boltzmann-statisztikával optimalizált statisztikai potenciális erőter felhasználásával. Az energiabeágyazás olyan matematikai modellt biztosít, amely tartalmazza mind a jellemző, mind a pozíciós információt, és amelyről megmutatjuk, hogy felhasználható transzfer tanuláshoz, hogy gyors, pontos és megbízható rendellenesség-előrejelzést érjünk el. Megközelítésünk felülmúlja a legtöbb jelenlegi módszert, és ígéretes utat kínál a javításhoz valamint a funkcionális rendezetlenség predikciójához is.

## **Mutációs mintázatok: biológia vs. informatika**

Németh Eszter, Szüts Dávid

*HUN-REN TTK Enzimológiai Intézet, Genomstabilitás Kutatócsoport*

A teljes genom szekvenálás forradalmasította a rákos daganatok genetikai hátterének kutatását. Fényt vetett arra, hogy a tumorigenezist vezérlő mutációk mellett több száz-több ezer további pontmutáció is detektálható a daganatsejtek genomjában. Bevett szokás ezeket a szekvencia környezet függvényében mutációs spektrumként ábrázolni. A daganatokban meghatározott mutációs spektrumok mögött ismétlődően előforduló mutációs mintázatokat azonosítottak, melyek különböző exogén vagy endogén mutagenikus hatásnak feleltethetőek meg.

Tetszőleges tumorminta mutációs spektruma modellezhető a referencia mutációs mintázatok lineáris kombinációjaként, feltárva a daganat hátterében álló biológiai folyamatokat. Könnyen belátható, hogy minél több mutációs mintázatot alkalmazunk, a modell annál tökéletesebb lesz. Ez a jelenség pedig túlllesztéshez vezethet a mintázatok alkalmazása, de már a származtatása során is. Ennek eredményeképpen a biológiailag értelmezhető mintázatok helyett azok töredékeit külön mintázatokként ismerjük el, megnehezítve a kutatási vagy diagnosztikai célú felhasználást.

Előadásomban három olyan példát mutatok be kutatásainkból, amely során tumor mutációs mintázatokat sejtmódellen végzett kísérletekkel igazolunk, és az eredmény alapján a mutációs mintázatok biológiai egységét kritikusán vizsgáljuk.

A kutatás a Nemzeti Kutatási, Fejlesztési és Innovációs Hivatal támogatásával (PD134818) valósult meg.

## **Transzformer és rekurrens architektúrát egyesítő hibrid neurális hálózat használata bakteriofágok és plazmidok detektálására**

Pethő Gergely

*DE*

A természetes nyelvek feldolgozásában 2017 óta szinte egyeduralmú transzformer architektúrájú neurális nyelvmodellek a DNS-szekvenciák feldolgozásában az utóbbi évekig nem kaptak számottevő figyelmet. A 2021-ben megjelent DNABERT modellre építve 2022-ben közzétették Bai és mtsai az INHERIT nevű kettős BERT nyelvmodellt, amelyet bakteriofágok felismerésére tanítottak be.

Előadásomban felhívom a figyelmet az INHERIT modell egyes a BERT architektúra jellegzetességeiből eredő gyengéire: 500 bp-nál tágabb kontextust képtelen figyelembe venni, a feldolgozás lassú. Rámutatok emellett a modell tanításának módjából adódó korlátaira: az INHERIT bináris osztályozó, amely valójában nem bakteriofágokat, hanem baktériumkromoszómákat ismer fel.

Ismertetem az általunk az INHERIT-re felépítve kifejlesztett InheRecurrent nevű, hibrid transzformer-rekurrens modellarchitektúrát, amelyet négy osztályból származó szekvenciák megkülönböztetésére tanítottunk be: 1. bakteriofág, 2. baktériumkromoszóma, 3. plazmid, 4. „egyéb”: eukarióták, archeák, nem fág vírusok. Modellünk a rekurrens komponensnek

köszönhetően kontextusukban látja az 500 bp-nyi szekvenciárszleteket, és így több tízezer bp terjedelmű szekvenciaszakaszokra is tud támaszkodni az osztályozás során. Amellett, hogy finomabb különbségtételre képes és így a gyakorlatban több helyzetben és több feladatra használható, pontossága felülmúlja az INHERIT-ét olyan feladatokban is, mint a baktériumok + plazmidok elkülönítése bakteriofágoktól, ha nincsenek a tesztadatok között „egyéb” osztályú szekvenciák.

## 2. szekció - Szerkezeti Bioinformatika

### A hidratáció szerepe a gyógyszertervezésben

Hetényi Csaba, Szél Viktor, Bayarsaikhan Bayartsetseg, Zsidó Balázs Zoltán, Börzsei Rita  
*PTE Farmakológiai és Farmakoterápiai Intézet, PTE Farmakoinformatikai Tanszék*

A vízmolekulák kulcsszerepet játszanak a legtöbb élettanilag fontos molekuláris kölcsönhatásban, beleértve a gyógyszer-célpont kölcsönhatásokat is. Szerepüket ezért a gyógyszerek tervezése során nem lehet figyelmen kívül hagyni. Ugyanakkor a kísérletes szerkezetmeghatározó eljárások számos korláttal küzdenek a vízmolekulák pozíciójának (a hidrátszerkezet) kimérésében. E korlátok meghaladásában a kutatók egyre nagyobb mértékben támaszkodnak az előadásban ismertetésre kerülő elméleti módszerekre. Bemutatásra kerül, hogy az elméleti úton szolgáltatott pontos vízpozíciók a tervezést mind szerkezeti mind energetikai (termodinamikai) oldalról jelentősen hatékonyabbá teszik.

### SH2db: online fehérjeszerkezeti adatbázis az SH2 doménekre

Bajusz Dávid<sup>1</sup>, Pándy-Szekeres Gáspár<sup>1</sup>, Takács Ágnes<sup>1</sup>, Elvin de Araujo<sup>2</sup>, Keserű György Miklós<sup>1</sup>

<sup>1</sup>*HUN-REN TTK Szerves Kémiai Intézet, Gyógyszerkémiai Kutatócsoport,*

<sup>2</sup>*University of Toronto Mississauga*

Az SH2 domén a foszfortirozin-alapú jelátvitel egyik kulcsfontosságú eleme: egy „olvasó” fehérjemodul, amely a foszfortirozint tartalmazó motívumok felismeréséért felelős. Mivel számos fehérjében előfordul, ezért rendkívül sokoldalúan szól bele a sejten belüli molekuláris szabályozásba és számos esetben az adott jelátviteli útvonalak onkogén mutáció miatt vagy egyéb okból történő meghibásodása esetén ígéretes gyógyszer-célpontként tekinthetünk rá. Térbeli szerkezete meglehetősen konzervált, különösen a központi motívumát jelentő, három béta-redőből álló struktúra megbízhatóan egymásra illeszthető az emberi szervezet 120 különféle SH2 doménje esetén is.

Az idén elkészült adatbázisunk (<http://sh2db.ttk.hu/>) az SH2 domén szerkezetek gyors kereshetőségét, egymásra illesztését, illetve különféle modellező, gyógyszertervező feladatok előkészítését hivatott elősegíteni a terület kutatói számára. Ehhez egy célzott adatbázis-struktúrát, térbeli illesztést, illetve generikus aminosav-számozási rendszert dolgoztunk ki, amelyek segítségével honlapunkon néhány kattintás segítségével érhetünk el releváns fehérjeszerkezeteket, vagy informatív összehasonlító ábrákat, munkameneteket.

### Fehérjék másodlagos szerkezetének jóslása - a szürke zóna

Lukács András

*ELTE TTK Matematikai Intézet, Mesterséges Intelligencia Kutatócsoport*

A legjobb módszerek másodlagos fehérje szerkezet megjóslására homológiákra épülnek. Amikor nem állnak rendelkezésünkre homológiák akkor fontossá válnak csak a vizsgált fehérje szekvenciájából következtető megoldások. Egy ilyen, LSTM neurális hálókra és semi-supervised tanításra építő modellt, Moffat és Jones (2021) S4PRED módszerét vizsgáltuk mély tanulási szempontból. Az a tapasztalat, hogy lecserélve az LSTM architektúrát 1D konvolúcióra és a tanítási paramétereket variálva az eredeti módszerhez hasonló teljesítményt lehet elérni, arra mutat, hogy közel járunk a tanítás során felhasznált adathalmazok kihasználásában.

### 3. szekció - Rendszerbiológia

#### **A metabolom evolúciójától a humán betegségekig**

Papp Balázs

*HUN-REN SzBK Biokémiai Intézet, HCEMM-BRC*

A metabolitok központi szerepet töltenek be a biokémiai reakcióhálózatokban, és koncentrációjuk megváltozása betegséget jelezhet. Az előadás központi hipotézise, hogy a metabolitkoncentrációk evolúciós konzervációja fontos információt hordoz a metabolom káros elváltozásairól, hasonlóan ahhoz, hogy a fehérjeszekvencia konzervációja információt ad a káros genetikai variánsokról. Az előadásban a metabolomok fajok közötti összehasonlítását felhasználva teszteljük e hipotézist. Eredményeink alapján az anyagcsere-betegségek biomarkerei pusztán az evolúciós konzerváltság alapján megkülönböztethetők más metabolitoktól, anélkül, hogy előzetes klinikai ismeretekre lenne szükség. Munkánk lehetővé teszi, hogy az emberi metabolomprofilok kiértékelésébe evolúciós információt is bevonjunk.

#### **Tumorsejtek gyógyszerérzékenységének predikciója - avagy hogyan vezethetnek félre az MI kiértékelési metrikák**

Szalai Bence

*Turbine Kft*

A tumorok gyógyszerérzékenységének megbízható predikciója alapvető fontosságú mind a személyre szabott orvoslás, mind az onkológiai gyógyszerfejlesztés szempontjából. A közelmúltban létrejött nagy léptékű tumor sejtvonala adatbázisok megkönnyítik a gépi tanulási módszerek alkalmazását e területen. E modellek lehetőséget teremthetnek szenzitivitási biomarkerek és új gyógyszer célpontok azonosítására. A modellek korrekt összehasonlításához elengedhetetlenek a jól kiválasztott kiértékelési metrikák és a standardizált "benchmark" adatsorok. Az előadás során bemutatjuk az általunk kidolgozott EFFECT (Evaluation Framework For predicting Efficiency of Cancer Treatment) keretrendszert a gyógyszerérzékenység predikciójára. A EFFECT segítségével standard gépi tanulási algoritmusok teljesítményét vizsgáltuk. Meglepő módon a legtöbb kiértékelési metrika nem tudott különbséget tenni az egyszerűbb és komplikáltabb modellek, illetve a biológiailag nem informatív és informatív adatokon tanított modellek között. E jelenség hátterében a gyógyszerérzékenységi adatsorokat általánosan torzító hatásokat azonosítottunk, és kidolgoztunk egy statisztikai módszert ("Bias Detector"), mely lehetővé teszi a gépi tanulási modelleknek a torzító hatásoktól független kiértékelését, ezáltal a valóban prediktív módszerek azonosítását.

#### **A hibaküszöb egy egyszerű, kompartmentalizált rendszerben**

Mikulecz Benjámin Áron

*ELTE, Biológia Doktori Iskola*

Az információfenntartás problémája központi szerepet játszik az élet eredetének kutatásában. A prebiotikus mutációs ráták olyan magasak voltak, hogy súlyos akadályt jelentettek az információhordozó szekvenciák hossza számára, ezzel korlátozva a fenntartható információ mennyiségét. Rövid szekvenciák közösségeit pedig a kompetitív kizárás és a közösség javait kihasználó, önző, parazita szekvenciák veszélyeztették. Azt a kritikus mutációs rátát, amely felett a szelekció nem képes fenntartani az információt, hibaküszöbnek nevezzük. A hibaküszöböt eredetileg determinisztikus, jól kevert rendszerben mutatták ki. Munkámban egy enzim-parazita populáció sztochasztikus modelljét vizsgálom explicit módon kompartmentekbe helyezve a replikátorokat a sztochasztikus korrektor modell mint keretrendszer alkalmazásával. A sztochasztikus korrektor modellt általában több, különböző sebességgel replikálódó enzimtípus kompetíciójának tanulmányozására alkalmazzák. Én itt egy enzimtípus és a nála potenciálisan gyorsabban replikálódó, parazita mutánsai evolúciós dinamikájának tanulmányozására használom. Szimulációkkal nyert adatokkal megmutatom, hogy az enzim a paraméterterem mely részén képes együtt élni a parazitákkal. A vizsgált paraméterek: (1) a kompartmentpopuláció



mérete, (2) az a replikátormennyiség, amelynél osztódik a kompartment, (3) a parazita replikációs rátája és (4) a kompartmentek fúziós rátája.

## **Sejtek transzkriptomikai korának meghatározása öregedési órák segítségével**

Zakar-Polyák Enikő, Kerepesi Csaba

*HUN-REN SZTAKI*

Az utóbbi években kiemelkedő figyelem irányul az öregedés folyamatának vizsgálatára és a belső mechanizmusainak feltárására. A legújabb eredmények azt mutatják, hogy omikai adathalmazokkal és a gépi tanulás eszközeivel lehetséges az egyének életkorának modellezése, úgynevezett "öregedési órák" segítségével, és így lehetőség nyílik az öregedési folyamat mérésére. Az öregedési órák többsége különböző (bulk) szövetmintákból származtatott epigenetikai információk alapján becsli az egyének életkorát, azonban a single-cell RNS szekvenálás egy újabb szintet nyithat az öregedés vizsgálatában.

Ebben a munkában több, mint egy millió emberi vérsejt gén expressziós profilja alapján hoztunk létre sejttípus-specifikus öregedési órákat, amelyekkel az egyes sejtek életkorát prediktáltuk. Több létrehozott óra is 8.5 és 10 év közötti átlagos abszolút hibával teljesít, valamint 0.4 és 0.5 közötti korrelációs együtthatóval a kronológiai és prediktált életkor között. Összehasonlításként sejttípus arányokon alapuló órákat is fejlesztettünk, valamint a két megközelítés kombinációjából eredő predikciókat is adtunk.

## **Hogyan segíthet a számítógépes sejtszimuláció a hatékonyabb gyógyszerfejlesztésben?**

Fichó Erzsébet

*Cytocast Hungary Kft*

A gyógyszeripar egyik legnagyobb kihívása a rendkívül hosszú és költséges fejlesztési folyamat. A gyógyszerjelöltek szűrése időigényes, és a pre-klinikai és klinikai szakaszokban bekövetkező kudarcok jelentős anyagi terheket rónak a gyógyszeriparra. A klinikai vizsgálatok során a kardiovaszkuláris gyógyszereknek mindössze 6.5%-a, az onkológiai készítmények pedig csupán 3.4%-a jut el a betegekhez. A klinikai hatékonyság hiánya mellett a legjelentősebb probléma a nemkívánatos mellékhatások gyakori jelenléte. ám még ezt az arányt is befolyásolhatják a nemkívánatos mellékhatások megjelenése.

A hagyományos fejlesztési módszerek alappillérei az in vitro és in vivo kísérletek. Azonban a számítógépes sejtsimulációval jelentős előrelépést ígér, pontosabb és részletesebb információk biztosításával.

A Cytocast Simulated Cell egy felhőalapú platform, amely lehetővé teszi a sejten belüli fehérje-fehérje kölcsönhatások, a kialakuló és megszűnő fehérjekomplexek modellezését. A platform különböző multi-omikai adatbázisok adatait felhasználva modellezi az egészséges sejtet, és megjósolja, hogy a különböző gyógyszeres kezelések milyen molekuláris változásokat indukálhatnak a sejtben. Az ágens-alapú sztochasztikus algoritmus továbbá képes előre jelezni az ezekhez a változásokhoz köthető mellékhatásokat különböző humán szövettípusban, ezáltal segítve a hatékony és biztonságosabb gyógyszerfejlesztést.

## 4. szekció - Orvosi bioinformatika

### Transzkriptomikai és fehérje-fehérje kölcsönhatás hálózatok elemzése gyógyszer-célpontok azonosítása céljából

Ágg Bence<sup>1,2</sup>, Benczik Bettina<sup>1,2</sup>, Balogh Olivér<sup>1</sup>, Bereczki Zoltán<sup>1</sup>, Váczy-Földi Máté<sup>1</sup>, Pétervári Máttyás<sup>1</sup>, Ferdinandy Péter<sup>1,2</sup>

<sup>1</sup>SE, Farmakológiai és Farmakoterápiás Intézet, HUN-REN-SE Rendszerfarmakológiai és Kardiometabolikus Kutatócsoport

<sup>2</sup>Pharmahungary Group

Háttér: Az iszkémiás szívbetegség kezelése céljából hipotézis vezérelt módszerekkel felderített terápiás megközelítések klinikai gyakorlatba történő átültetése mindezedig sikertelen volt. A probléma hátterében az elfogult kísérlettervezést, illetve a komorbiditások módosító hatásának figyelmen kívül hagyását feltételezhetjük.

Célkitűzések: Célul tűztük ki ezért olyan szoftverek kifejlesztését és alkalmazását, melyek lehetővé teszik a hipotézismentes megközelítések alapját képező, nagy áteresztőképességű molekuláris biológiai módszerekkel nyert adathalmazok hálózatelméleti kiértékelése révén új gyógyszer-célpontok felderítését.

Módszerek és eredmények: Az interaktómok hipotézismentes megközelítések alkalmazását hátráltató hiányosságainak kezelése céljából a fehérje-fehérje kölcsönhatás hálózatok elemzésére alkalmas feltételes generatív adverzális hálózat (cGAN) alapú szoftvert ([https://github.com/semmelweis-pharmacology/ppi\\_pred](https://github.com/semmelweis-pharmacology/ppi_pred)) fejlesztettünk. Szoftverünk validálása az International Network Medicine Consortium által szervezett együttműködés keretében a STRING és a BioGRID adatbázisok fehérje-fehérje kölcsönhatás hálózatain, illetve yeast-two-hybrid módszerrel történt. 26 módszer teljesítményének szisztematikus összehasonlítása során cGAN alapú szoftverünk az élmezőnybe került. A kis nem kódoló RNS-ek, azon belül is a mikroRNS-ek különféle megbetegedések pathomechanizmusában betöltött szerepének vizsgálatára mikroRNS-target hálózatok integratív elemzését megvalósító, felhasználóbarát webes felületen keresztül is elérhető, miRNAtarget™ nevű (Pharmahungary, Szeged, <https://mirnarget.com>) szoftvert hoztunk létre. A prediktált és kísérletesen validált mikroRNS-target kölcsönhatás adatbázisokra (microrna.org, miRDB, miRTarBase) támaszkodó miRNAtarget szoftver által azonosított targetek mRNS- és fehérjeszintű validációja sikerrel megtörtént számos, többek között kardiovaszkuláris kórképeknek és azok komorbid kórállapotainak vizsgálatára szolgáló kísérleti modell esetében.

Következtetések: Elsőként sikerült igazolnunk, hogy egy kizárólag a hálózattopológia elemzésére szorítkozó, így a fehérjék molekuláris tulajdonságainak ismerete hiányában is működőképes, cGAN alapú gépi tanulási algoritmus alkalmas fehérje-fehérje kölcsönhatások hatékony predikciójára. Emellett a költséghatékonyan meghatározható kis RNS transzkriptomikai profilok elemzésére szolgáló, sikeresen validált miRNAtarget szoftverünkkel több potenciális gyógyszer-célpontot azonosítottunk.

### A kissejtes tüdőrák genomikai jellemzése

Pongor S. Lőrinc

HCEMM Onkogenomikai és Epigenetikai Kutatócsoport

A kissejtes tüdőrák (SCLC) a tüdőrákok egy igen agresszív formája, amelynél gyakran már diagnóziskor áttétek találhatók. Jelenleg az SCLC kutatás célja megérteni a genom-szintű genetikai és epigenetikai dinamikus változásokat, amik a heterogenitást és a gyógyszerrezisztenciát irányítják. Korábbi munkánkban sikerült bemutatni, hogy az SCLC heterogenitását hajtó MYC paralóg amplifikációk főként az extrakromoszomális DNS (ecDNS) elemeinek részét képezik, illetve jelenlétük rosszabb túlélési prognózissal jár együtt. A DNS és epigenetikai szekvenálási adatok integrálásával célunk az SCLC altípusaira jellemző transzkripció faktorok által vezérelt genom lókuszek azonosítása, amelyek befolyásolják a gén transzkripciót. A transzkripció profil alapján új gyógyszer célpontokat lehet azonosítani, amik segíthetnek az SCLC kezelésének a fejlesztésében.

## **Vesetumorok célzott terápiás célpontjainak kifejeződése egysejt RNS-szekvenálási adatok alapján**

Dobolyi Zsófia<sup>1</sup>, Menyhárt Otília<sup>1,2</sup>, Györffy Balázs<sup>1,2</sup>

<sup>1</sup>SE, Bioinformatika Tanszék

<sup>2</sup>HUN-REN, TTK, Onkológiai Biomarker Kutatócsoport

A vesesejtes karcinóma (RCC) világszerte átlagosan 180 000 ember halálát okozza. Ez a vesét érintő leggyakoribb rákos megbetegedés, azonban kezelése nehéz, mivel gyakran alakul ki terápiás rezisztencia. Feltételezésünk szerint ennek oka a célzott gének mérsékelt expressziója lehet a rákos sejtekben. Kutatásom során 174 611 sejt expressziós profilját vizsgáltam szabadon hozzáférhető egysejt RNS-szekvenálási adatbázisok integrálásával, az analízishez bioinformatikai munkafolyamatot terveztem Python programozási nyelven. A CellMarker2.0 és a CancerSEA adatbázisok markergénjeit használtam az integrált adathalmaz különböző sejtípusainak annotálásához. A DrugBank adatbázis segítségével összegyűjtöttem a jelenleg forgalomban lévő-, vagy klinikai vizsgálatokban tesztelt RCC elleni gyógyszereket, és azonosítottam ezen gyógyszerek célgénjeit. A legtöbb rendszeresen célzott gént csak kis arányban expresszálták az annotált tumorsejtek: PARP - 4,95% és mTOR - 2,51%. Míg a VEGFA, a jelenleg leghatékonyabb gyógyszerek célpontja az annotált tumorsejtek 77,29%-ában fejeződik ki. Az egysejt-szintű génexpressziós profilok tehát korrelálhatnak az ismert célzott terápiák átlagos hatékonyságával. Továbbá olyan géneket is azonosítottam, amelyek az annotált tumorsejtek több, mint felében expresszálódnak. Ezek az új potenciális terápiás célgének a következők: UGT2B7; CA9; és VEGFB. Ezekre a célpontokra érdemes gyógyszerfejlesztést kezdeményezni.

## **Arc fotó alapú biológiai óra elemzése mortalitásra és foglalkozásokra**

Fejes Iván

HUN-REN Sztaki

Az emberi arc rendkívül sok, az öregedést jelző mintázatot tartalmaz, kérdéses, hogy ezekből a mintázatokból milyen módszerekkel és milyen pontossággal lehet kronológiai és biológiai életkort, esetleg mortalitást becsülni.

A SZTAKI-nál fejlesztettünk egy arc fotó-életkor biológiai órát mely 6.5 év átlaghibával képes kronológiai életkort becsülni. Kutatásom során az arc fotó-életkor predikciók biológiai aspektusait vizsgáltam, miszerint lehet-e gyorsult, illetve lassult öregedésre következtetni az arcképek alapján, esetleg összefügg-e különböző életviteli változókkal a modellünk becsléseinek eredménye.

A kutatásom folyamán az IMDB-WIKI adathalmazt használtam fel, mely híres emberek képeit tartalmazza. Az adathalmazban található személyekről életviteli adatokat, foglalkozásukat, halálkozásuk időpontját és módját, betegségeiket a WIKIDATA publikus adatbázisból nyertem ki. Az adathalmaz képeire a modellünk becslései alapján age accelerationt (öregedés gyorsultsági faktort) számoltunk ki, majd személyenként csoportosítva vizsgáltam az age acceleration és az életviteli adatok közötti összefüggéseket.

Eredmények: Kijött, hogy 45 év felett szignifikánsan korrelál az age acceleration és a halálig eltelt idő és az alacsonyabb age acceleration értékkel rendelkezők tovább élnek, mint a magasabb értékkel rendelkezők. Szintén hasonlóan szignifikáns eredmény jött ki, amikor Kaplan-Meier túlélési analízist végeztem gyorsult és lassult öregedők csoportjai között. 45 éves korcsoport felett 5-10 évvel később feleződött le a lassult öregedők populációja, a gyorsult öregedők megfeleződéséhez képest. Kijött még hogy a foglalkozások közül a sportolók öregednek a leglassabban és a hosszantartó ülőmunkát végzők a leggyorsabban. Az eredmények nagyrésze hasonlít az epigenetikai órákkal történt kutatások eredményeire, ezzel igazolható, hogy modellünk képes a biológiai életkor becslésére.

## **Kvantitatív szerológia növekedési függvények alkalmazásával**

Prechl József<sup>1</sup>, Kovács Ágnes<sup>2,3</sup>, Papp Krisztián<sup>1</sup>, Hérincs Zoltán<sup>1</sup>, Pfeil Tamás<sup>3</sup>

<sup>1</sup>Diagnosticum Zrt., K+F Laboratórium,

<sup>2</sup>Állatorvostudományi Egyetem, Biostatistika Tanszék

<sup>3</sup>ELTE, Alkalmazott Analízis és Számításmatematikai Tanszék

A vérben található specifikus ellenanyagok kimutatása fontos diagnosztikai eszköz, azonban azok kvantitatív jellemzése máig megoldatlan probléma. A gondot az jelenti, hogy nagyon sok ismeretlen van a képletben: sok limfocita klón vesz részt az immunválaszban, ennek megfelelően sok ellenanyag szerkezet, változó affinitásokkal és koncentrációkkal járul hozzá a mérhető reakciókhoz. Egy mikromátrixokon alapuló technológiai megoldást ötvözve az eredmények mélyebb matematikai elemzésével, Richards növekedési függvények alkalmazásával, valós biokémiai mértékegységeket eredményező technológiát dolgoztunk ki. A módszert autoimmun (cöliákia) és fertőzőes (SARS-CoV-2) folyamatok marker ellenanyagainak kimutatására is alkalmaztuk. A technológia kiegészítheti a jelenlegi orvosdiagnosztikai módszereket és utat nyithat a kvantitatív immunbiológia irányába.

